BASIC COMPLEX ANALYSIS OF ONE VARIABLE

Anant R. Shastri
Department of Mathematics
Indian Institute of Technology, Bombay

And the detailed exposition can no less obfuscate than the overly terse. — **J.Kepler**

Under construction

Preface

Every mathematics student has to learn complex analysis. In fact every mathematics teacher should teach a course in complex analysis at least once. However, every mathematics teacher need not write a book on complex analysis too. Nevertheless, here is yet another book on this topic and I offer no justification for it.

This book is intended as a text/reference book for a first course in complex analysis (of duration one year or two semesters) for M. Sc. students in Indian universities and institutes of technologies. It is suitable for students who have learnt to deal with basic set theoretic and $\epsilon - \delta$ arguments. I assume that the student has been exposed to basic differential and integral calculus of one real variable. It is also desirable that the student is exposed to some calculus of two variables, though, strictly speaking this is not necessary. I hope that as the course proceeds, the student acquires more and more sophistication.

This book is the outcome of the lecture notes for the courses that I have taught at our department to M. Sc. students. In our department we teach roughly the material in the first seven chapters in the first semester with three hours of lecture and one hour tutorial per week, as a compulsory course. The last three chapters can be covered in a leisurely fashion in a semester with three hours a week, as an elective course. However, it is also possible to adopt a slightly different order of presenting this material. For instance, material in chapter 3, can be postponed until you come to chapter 9. On the other hand, most of the material in chapter 8 can be covered immediately after chapter 4.

I have tried to keep a dialogue style as far as possible. Throughout the book, I have tried to remember my own difficulties as a student. A student who wants to learn from this book should try to answer the questions that are being asked from time to time and then proceed. Also, she should pause and reflect every time phrases such as 'it is obvious', or 'it follows' are used, to see whether this really is so. For instance, when I am saying that something is obvious and it is not at all obvious to her, she should perhaps re-read the material just before or some relevant topic that has been covered before. Generally, comments which are inside square brackets are meant for students who are above average or those who have a better background. The subsequent materials do not depend upon them. Enough exercises have been included to take care of students of various calibre. Some of them have been marked with a star, not to discourage the student from trying it but to tell her that even if she does not get it at the first attempt, it is alright. Of course some of them may need several attempts and the students may not have so much time to devote. In any case, hints and solutions are given to almost all exercises, so that the student can compare her answers with them.

There cannot be anything new to say in such a widely used elementary topic. I have freely borrowed materials from several standard books, a bibliography of which has been included. There are many other books worth mentioning as good books but I have not borrowed anything from them. For instance, the choice of topics is almost a subset of those in Ahlfors's book, which has influenced me to a great extent. Three other books that I really liked are [P-L], [S] and [R]. I wrote down the material in section 3.5, and later on was very happy to read similar exposition in Remmert's book, the English addition of which had come to our library just then.

During the past five years, ever since I latexed my lecture notes, I have received impetus and help from several quarters in converting these notes into a book. It all began with the typical query by several of my students: *'Sir, are you going to write a book?.....why not?'* Prof. M. G. Nadkarni, Prof. M. S. Narasimhan, Prof. C. S. Seshadri, Prof. K. Varadarajan etc. have put in a lot of encouraging words. My wife Parvati was first to go through those primitive notes and gave me a vague idea of the kind of task before me. Prof. R. R. Simha and Prof. R. C. Cowsik have offered many valuable comments apart from encouragement. Indeed, apart from providing troubleshooting suggestions with latex, Cowsik went through the pre-final version and has corrected several mathematical , grammatical and typographical errors. (The author is solely responsible for whatever errors still persist, despite this.) I am indebted to all these people and many more.

During the summer of 1995, I spent three months at Inter-national Centre for Theoretical Physics, Trieste, out of which about a month I spent enlarging and polishing the notes. Because of the excellent facilities and environment there and the free time one gets, I could do a lot in that single month. Above all, it was quite enjoyable. I would like to mention that the final conversion of the lecture notes into the book form was

carried out under the Curricular Development Programme of my institute.

Preface to the II-edition

It gives me great pleasure to place this thoroughly revised edition of my book, though somewhat belatedly.

This edition contains more than 470 pages as compared to 300 pages in the first edition. Besides bringing further clarity in the presentation by reorganizing the material, I have added quite a bit of new material such as the homotopy version of Cauchy's theorem, Runge's theorem and a whole chapter on periodic functions culminating into proof of Picard theorems. Throughout I have added more exercises also. In few places I have cut down some material as well.

Other things I have said in the preface to the first edition is valid for this edition as well. This edition contains enough material for a first course (one year or equivalently two semesters) in complex analysis at M. Sc. level at Indian universities and institutions. One of the new features of this edition is that part of the book can be fruitfully used for a semester course for Engineering students, who have a good calculus background. Look at the dependence tree to decide your route. I have myself followed :

(chapter.section) 1.1-1.17; 2.1-2.4; 3.1, 3.5, 3.6, 3.7; 4.1, 4.2, 4.4-4.10; 5.1-5.7; 6.1-6.6.

My sincere thanks are due to many colleagues, friends and students who have offered comments which has helped in improving this edition. I should especially mention R. R. Simha, S. S. Bhoosnurmath, Gowri Navada, Goutam Mukherjee who have meticulously gone through one or the other version of this manuscript listing out typos, raising objections and offering suggestions. I had several opportunities to teach this material at the ATM Schools of National Board for Higher education organized by various persons at various locations such as Bhaskaracharya Pratisthan, DEparment of Mathematics Delhi University etc. I have benefited from interaction with college teachers and research students who participated in these schools.

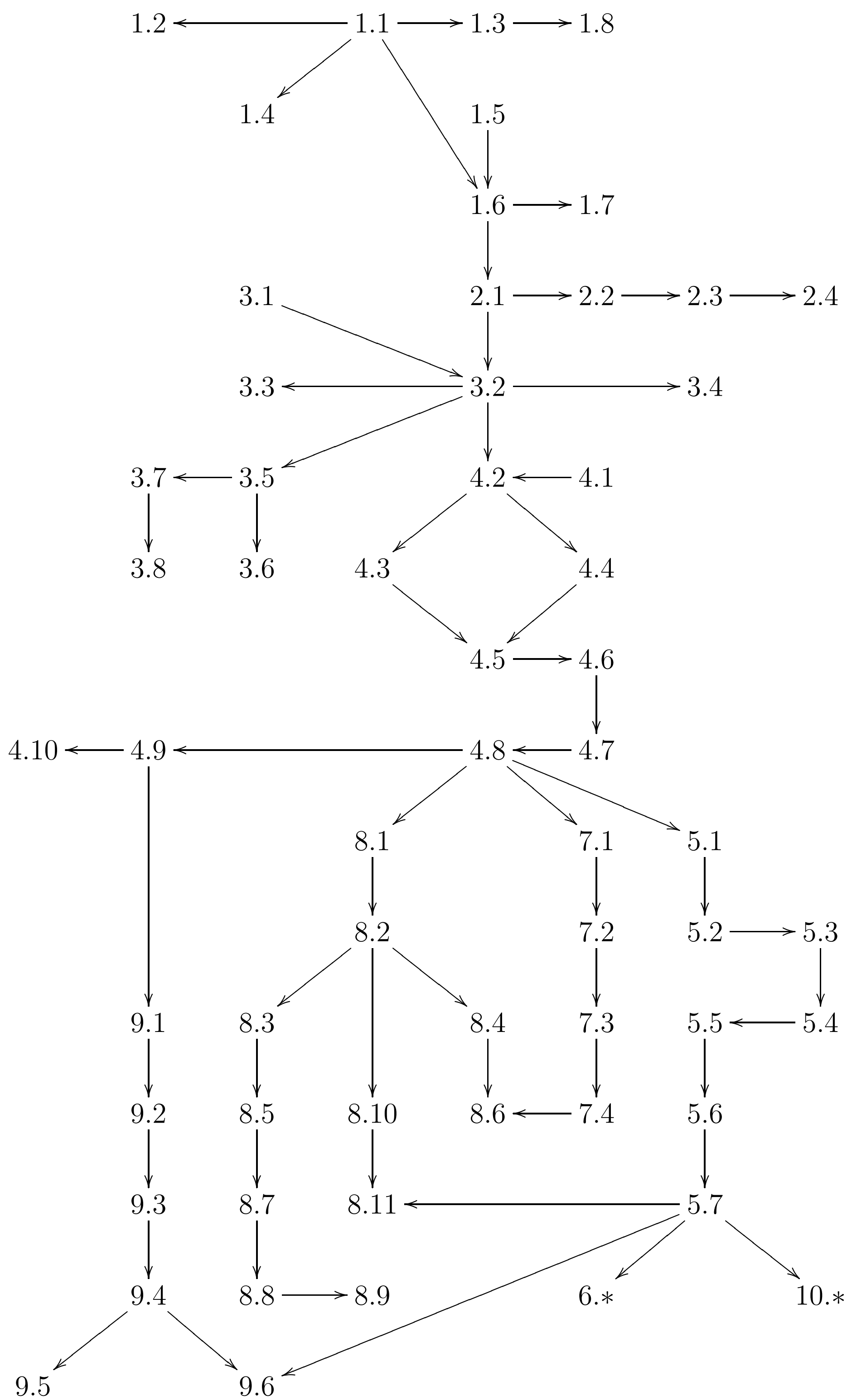
However, I am solely responsible for whatever inaccuracies still persist and will happily receive reports of any such and promise pos the corrections on my website. The process of revision had started right from the day I received the author's copies of the I edition. It is said that you never finish writing a book but you abandon it at some stage. The same seems to apply to revising a book as well.

A. R. Shastri

Spring, 2010

Contents

Dependence Tree



List of Symbols

Symbols	Section no.	Symbols	Section no.
$\mathbb{R}$	1.1	$\mathbb{R}^2$	1.1
$\mathbb{C}$	1.1	$\Re(z)$	1.1
$\Im(z)$	1.1	$\mathbb{N}$	1.2
$\mathbb{R}^n$	1.5	$B_r(\mathbf{x})$	1.5
$\overline{B}_r(\mathbf{x})$	1.5	$\mathbb{S}^1$	1.8
$\mathbb{C}[z]$	2.2	$\mathbb{C}(z)$	2.2
$\mathbb{K}$	2.3	$\mathbb{K}[[t]]$,	2.3
exp	2.4	$\mathcal{C}^r$	3.2
$\mathcal{C}^\infty$	3.2	$\dfrac{\partial}{\partial z}, \dfrac{\partial}{\partial \bar{z}}$	3.4
∇^2	3.4	$\boldsymbol{H}$	3.6
$GL(2, \mathbb{C})$	3.7	$\mathbb{D}$	3.7
$\mathbb{S}^2$	3.8	$\widehat{\mathbb{C}}$	3.8
$\mathbf{P^1}(\mathbb{C})$.	3.3	$\mathcal{H}(\Omega)$	5.1
$\mathcal{M}(\Omega)$	5.2	$R_a(f) = Res_a(f)$	5.3
$A(a; r_1, r_2)$	5.3	$\eta(\gamma, z_0)$	5.5
$\mathcal{L}(f)$	6.5	$\sum_{-\infty}^{\infty} e$	8.2
$\wp$	11.3	$\sum_{\gamma}'$	11.5

Chapter 1

Basic Properties of Complex Numbers

1.1 Arithmetic of Complex Numbers

We shall denote the set of real numbers by $\mathbb{R}$. The set of all ordered pairs (x, y) of real numbers x, y will be denoted by $\mathbb{R}^2$. Recall that two ordered pairs $(x_1, y_1), (x_2, y_2)$ are equal to each other iff $x_1 = x_2$, and $y_1 = y_2$.

On $\mathbb{R}^2$, we define the operation of addition and multiplication by the rules:

$$\boxed{\begin{array}{rcl} (x_1, y_1) + (x_2, y_2) & = & (x_1 + x_2, y_1 + y_2) \\ (x_1, y_2)(x_2, y_2) & = & (x_1x_2 - y_1y_2, x_1y_2 + y_1x_2). \end{array}} \tag{1.1}$$

We call the set of all ordered pairs of real numbers together with these two arithmetic operations the set of '*complex numbers*' and denote it by $\mathbb{C}$. Thus, a complex number is a member of $\mathbb{C}$. They will be denoted by symbols such as z, w etc.. If $z = (x, y)$ is a complex number then

$$\boxed{\Re(z) := x; \quad \Im(z) := y} \tag{1.2}$$

are respectively called the *real part* and *imaginary part* of z. Verify that this addition and multiplication in $\mathbb{C}$ follow various rules of arithmetic such as:

(Ai) $(x_1, y_1) + (x_2, y_2) = (x_2, y_2) + (x_1, y_1); \quad (x_1, y_1)(x_2, y_2) = (x_2, y_2)(x_1, y_1)$.

(Aii) $[(x_1, y_1) + (x_2, y_2)] + (x_3, y_3) = (x_1, y_1) + [(x_2, y_2)] + (x_3, y_3)]$.
(Aiii) $(x_1, y_1)[(x_2, y_2) + (x_3, y_3)] = (x_1, y_1)(x_2, y_2) + (x_1, y_1)(x_3, y_3)$.
(Aiv) $(0, 0) + (x, y) = (x, y); \quad (1, 0)(x, y) = (x, y)$.
(Av) $(x, y) + (-x, -y) = (0, 0)$.

In particular, verify that

$$(0, 1)^2 = (0, 1)(0, 1) = (-1, 0). \tag{1.3}$$

We can define an operation of $\mathbb{R}$ on $\mathbb{R}^2$ as follows:

$$r(x, y) = (rx, ry) \tag{1.4}$$

Verify this operation satisfies the usual properties:
(V1) $(r + s)(x, y) = r(x, y) + s(x, y); \quad rs(x, y) = r(s(x, y))$.
(V2) $r[(x_1, y_1) + (x_2, y_2)] = r(x_1, y_1) + r(x_2, y_2)$.
(V3) $1(x, y) = (x, y)$.

Using this operation, we can write

$$(x, y) = x(1, 0) + y(0, 1) \tag{1.5}$$

in a unique way. Next, we observe that the assignment $x \mapsto (x, 0)$ defines a one-one mapping of the real numbers onto all the complex numbers with their imaginary part equal to zero. Moreover this assignment preserves the arithmetic operations on either side. Via this map, we can identify the set of all complex numbers having their imaginary part zero with the set of real numbers.

The property (1.3) has a special significance now. We know that -1 does not have any square-root inside the real number system. If we use complex numbers we get two such roots of the equation $z^2 = -1$, viz., $\pm(0, 1)$. In view of (1.5), we shall cook up a notation $\imath$ to denote $(0, 1)$. Now (1.5) can be restated as follows: **every complex number can be expressed in a unique way as**

$$(x, y) = x + \imath y. \tag{1.6}$$

We need not ignore the other square-root of -1 viz., $-(0, 1)$. Indeed, the choice of denoting $(0, 1)$ as $\imath$ was totally arbitrary and we could have chosen $-(0, 1)$ to play this role. The symmetry involved in this phenomenon can be expressed in an elegant fashion: we want to have a mapping $\phi : \mathbb{C} \longrightarrow \mathbb{C}$ which send $\imath$ to $-\imath$ and does not disturb the real numbers $(x, 0) \in \mathbb{C}$. We expect ϕ to respect the properties (V1)-(V3). It turns out that there is such a map and we have

$$\phi(a + \imath b) := a - \imath b.$$

We shall have simpler and special notation as well as a name for this very important map.

Definition 1.1.1 We define the *complex conjugate* of a complex number $z = x + \imath y$ by

$$\bar{z} := x - \imath y. \tag{1.7}$$

Observe that $\Re(z)$ and $\Im(z)$ can be expressed in terms of z and $\overline{z}$ by the formula

$$\boxed{\Re(z) = \frac{z + \overline{z}}{2}; \qquad \Im(z) = \frac{z - \overline{z}}{2\imath}.} \tag{1.8}$$

Verify that

$$\boxed{\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \quad \overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}, \quad \overline{\overline{z}} = z.} \tag{1.9}$$

Moreover, verify that $\overline{z} = z$ iff z is real.

Definition 1.1.2 Given $z \in \mathbb{C}, \ \ z = a + \imath b$, we define its *absolute value* (*length*) $|z|$ to be the non-negative square root of $a^2 + b^2$, i.e.,

$$|z| := \sqrt{(a^2 + b^2)}.$$

Remark 1.1.1 $|z|^2 = z\overline{z}$. Hence for any $z \in \mathbb{C}, \ \ |z| \neq 0 \Longleftrightarrow z \neq 0$ and in that case $z^{-1} = \overline{z}|z|^{-2}$.

Now, let us list some of the easy consequences of the definitions that we have made so far. As usual, z, z_1, z_2 etc. denote complex numbers.

Basic Identities and Inequalities

(B1) $|\overline{z}| = |z|$.

(B2) $|z_1 z_2| = |z_1||z_2|$.

(B3) $|\Re(z)| \leq |z|$ (resp. $|\Im(z)| \leq |z|$); equality holds iff $\Im(z) = 0$ (resp. $\Re(z) = 0$).

(B4) *Cosine Rule:* $|z_1 + z_2|^2 = |z_1|^2 + |z_2|^2 + 2\Re(z_1\overline{z_2})$.

(B5) *Parallelogram Law :* $|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2(|z_1|^2 + |z_2|^2)$.

(B6) *Triangle inequality :* $|z_1 + z_2| \le |z_1| + |z_2|$ and equality holds iff one of the z_j is a non-negative multiple of the other.

These properties will be used quite often in what follows. So, the reader should acquire a thorough familiarity with them as soon as possible. Try to write down the proofs of each statement below by yourself before reading out the proof and compare your proof with the given proof.

Proof: The proofs of (B1), (B2), and (B3) are easy. In (B4) we have,

$$\begin{aligned} |z_1 + z_2|^2 &= (z_1 + z_2)(\overline{z_1} + \overline{z_2}) \\ &= |z_1|^2 + |z_2|^2 + (z_1\overline{z_2} + z_2\overline{z_1}) \\ &= |z_1|^2 + |z_2|^2 + 2\Re(z_1\overline{z_2}). \end{aligned}$$

To prove (B5), replace z_2 by $-z_2$ in (4) to obtain

$$|z_1 - z_2|^2 = |z_1|^2 + |z_2|^2 - 2\Re(z_1\overline{z_2}).$$

Now add this to (B4) to obtain (B5).

To prove (B6), it is enough to show that $|z_1 + z_2|^2 \le (|z_1| + |z_2|)^2$.
But from (B4),(B3), (B2) and (B1) we have,

$$\begin{aligned} L.H.S. &= |z_1|^2 + |z_2|^2 + 2\Re(z_1\overline{z_2}) && \text{from (B4)} \\ &\le |z_1|^2 + |z_2|^2 + 2|z_1\overline{z_2}| && \text{from (B3)} \\ &= |z_1|^2 + |z_2|^2 + 2|z_1||z_2| \text{ from (B2) (B1)} \\ &= R.H.S. \end{aligned}$$

Equality holds iff $\Re(z_1\overline{z_2}) = |z_1\overline{z_2}|$, i.e., $z_1\overline{z_2} = t$ is real and non-negative. If this number is 0 then either $z_1 = 0$ or $z_2 = 0$. So assume $t \ne 0$. Then $tz_2 = z_1z_2\overline{z_2}$. Therefore, $z_2 = z_1|z_2|^2/t$, as required.

Remark 1.1.2 To see the justification for the names of the properties (B4), (B5) and (B6) you will have to wait for a while. Property (B4) seems to be the most fundamental amongst them at least for the simple reason that it was used in proving each one of the others. Also, there are several situations in which we have to use (B4) directly rather than (B5) or (B6). A number of exercises below and at the end of the next section illustrate this point.

Exercise 1.1

1. Express the following quantities in the form $x + \imath y$.
   (a) $(1+2\imath)^2$, $\dfrac{3}{3+4\imath}$, $(1+\imath)^4 + (1-\imath)^4$;
   (b) $(a+\imath b)^4$, $(a+\imath b)^{-2}$, $(a^2+\imath b^2-1)/(a^2+\imath b^2+1)$, $(a, b \in \mathbb{R})$;
   (c) $\left(\dfrac{-1 \pm \imath\sqrt{3}}{2}\right)^3$, $\left(\dfrac{1 \pm \imath\sqrt{3}}{2}\right)^6$, $\dfrac{(1+\imath)^7}{(1-\imath)^7}$, $\dfrac{1+\imath\tan\alpha}{1-\imath\tan\alpha}, \alpha \in \mathbb{R}$.

2. Rewrite the arithmetic rules (Ai)–(Av) in this section in the notation $x + \imath y$ for complex numbers.

3. If $z = x + \imath y$, show that $\dfrac{|x|+|y|}{\sqrt{2}} \le |z| \le |x| + |y|$.

4. Prove that $|\,|z_1| - |z_2|\,| \le |z_1 - z_2|, \quad z_1, z_2 \in \mathbb{C}$.

5. Prove that for any two complex numbers z_1, z_2 we have:
   (a) $|1 - \bar{z}_1 z_2|^2 - |z_1 - z_2|^2 = (1-|z_1|^2)(1-|z_2|^2)$.
   (b) $|z_1 + z_2| = \left|\dfrac{|z_1|}{|z_2|} z_2 + \dfrac{|z_2|}{|z_1|} z_1\right|, \quad z_1, z_2 \ne 0$.

6. If $|z| = 1$, compute $|1+z|^2 + |1-z|^2$.

7. Show that $(|z_1| + |z_2|)\left|\dfrac{z_1}{|z_1|} + \dfrac{z_2}{|z_2|}\right| \le 2|z_1 + z_2| \quad z_1, z_2 \ne 0$.

8. Prove the following identities:
   (a) $\displaystyle\sum_{1 \le r < s \le n} |z_r + z_s|^2 = \left|\sum_{r=1}^n z_r\right|^2 + (n-2)\sum_{r=1}^n |z_r|^2$;
   (b) $\displaystyle\sum_{1 \le r < s \le n} |z_r - z_s|^2 + \left|\sum_{r=1}^n z_r\right|^2 = n\sum_{r=1}^n |z_r|^2$.

9. (a) Prove the following **generalized cosine rule:**

$$\left|\sum_{j=1}^n \alpha_j\right|^2 = \sum_{j=1}^n |\alpha_j|^2 + 2\Re\left(\sum_{1 \le j < k \le n} \alpha_j \overline{\alpha_k}\right). \tag{1.10}$$

   (b) Substitute $\alpha_j = z_j w_j$ and simplify to obtain the following **Lagrange's identity:**[1]

[1]Joseph Louis Lagrange (1736-1813) was born in Turin to Italian-French parents. He was an analyst from beginning to end. His main contribution is in unifying mechanics through his calculus of variations.

$$\boxed{\left|\sum_{j=1}^{n} z_j w_j\right|^2 = \left(\sum_{j=1}^{n} |z_j|^2\right)\left(\sum_{j=1}^{n} |w_j|^2\right) - \sum_{1\le j<k\le n} |z_j\overline{w_k} - z_k\overline{w_j}|^2.} \tag{1.11}$$

(c) Deduce *Cauchy's*[2] *Inequality :*

$$\boxed{\left|\sum_{j=1}^{n} z_j w_j\right|^2 \le \left(\sum_{j=1}^{n} |z_j|^2\right)\left(\sum_{j=1}^{n} |w_j|^2\right).} \tag{1.12}$$

(d) Show that equality occurs in (1.12) iff the ratios $z_j/\overline{w_j}$ are the same for all $j = 1, 2, \ldots, n$.

10. Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$ with $a_j \in \mathbb{R}$. Show that $p(\overline{z}) = \overline{p(z)}$. Conclude that if z is a root of a polynomial p with real coefficients then so is $\overline{z}$.

11. Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0,\ a_n \neq 0, n \geq 1$ be a polynomial of a complex variable with complex coefficients. Show that $|p(z)| \to \infty$ as $|z| \to \infty$.

12. Discuss $\lim\limits_{z\to\infty} \dfrac{p(z)}{q(z)}$, where p, q are some given polynomials.

13. Using the fact that every positive real number has a square-root, first show that $z^2 = w$ can be solved in $\mathbb{C}$ for any real number w. Then show that we can solve this even for any complex number w. Next show that every quadratic equation with complex coefficients has a solution in $\mathbb{C}$. Finally show that $z^{2^n} = w$ has a solution for any complex number w. [Hint: See next section.]

1.2 *Why the Name Complex

As a student, I wondered about the strange way the multiplication of complex numbers is defined in (1.1) as compared to the natural way the addition is defined therein. I thought *'Certainly the co-ordinatewise multiplication*

$$(x_1, y_1)(x_2, y_2) = (x_1 x_2, y_1 y_2) \tag{1.13}$$

[2]Augustin Louis Cauchy (1789-1857) was a French mathematician, an engineer by training. He did pioneering work in analysis and the theory of permutation groups, infinite series, differential equations, determinants, probability and mathematical physics.

is a natural choice! Why make life complicated?' If such questions do not bother you at this moment, you may skip reading this section and proceed further and come back to it if and when you feel.

In this section, I shall try to explain why I was wrong. Indeed, one of the objects of this book is to explain a number of such questions which bothered me as a student. So, even if you are quite familiar with the complex numbers, I advise that you should at least browse through each section, before proceeding further.

In the previous section, implicitly, we assumed that the reader is familiar with the arithmetic of real numbers.

Let us recall a few of these things here a little more systematically. The addition and multiplication of real numbers obey the following rules:

FI *The law of commutativity*: $a + b = b + a$; $ab = ba$, for all $a, b \in \mathbb{R}$.
FII *The law of associativity*: $(a + b) + c = a + (b + c)$; $(ab)c = a(bc)$, for all $a, b, c \in \mathbb{R}$.
FIII *The law of distributivity*: $(a + b)c = ac + bc$, for all $a, b, c \in \mathbb{R}$.
FIV *The law of identity*: $a + 0 = a$; $a1 = a$, for all $a \in \mathbb{R}$.
FV *The law of additive inverse*: Given any $a \in \mathbb{R}$, there exists a unique $x \in \mathbb{R}$ such that $a + x = 0$.
FVI *The law of multiplicative inverse*: Given $a \in \mathbb{R}, a \neq 0$, there exists a unique $x \in \mathbb{R}$ such that $ax = 1$.

It is very convenient to refer to all these properties by a single name and the name here is *'field'*. Thus one says the set of all real numbers forms a field. The idea is that any property of real numbers that is a mere logical consequence of the above set of properties FI-FVI will hold for other systems which are also fields. Indeed, there are many other fields which are of great interest in mathematics.

The essence of numbers is in their property that any two of them can be 'compared'. Mathematically, we express this by saying that there is a *total ordering* '$<$' on the set $\mathbb{R}$. More precisely,

FVII Given any two real numbers a, b, either $a = b$ or $a < b$ or $b < a$.
This ordering has an intimate relation with the two arithmetic operations. Once again, mathematically, we express this by saying that the order '$<$' is compatible with the above arithmetic operations, viz.,
FVIII $a < b \Longrightarrow a + c < b + c$ and $ad < bd$ for all $a, b, c \in \mathbb{R}$ and $d > 0$.[3]

[3] A field with a total order satisfying FVIII is called an ordered field.

The *Analysis* on real numbers begins with the realization that $\mathbb{R}$ is *complete* with respect to this ordering, viz.,

$$\boxed{\begin{array}{c}\texttt{Every non empty subset of } \mathbb{R} \texttt{ which is bounded above}\\ \texttt{has a least upper bound in } \mathbb{R}.\end{array}} \tag{1.14}$$

This is also called the *least upper bound property.*

So far as this course is concerned, you may simply take all the above properties as part of the definition of real numbers.

Recall a real number $r \in \mathbb{R}$ is said to be positive (resp. negative) if $r > 0$ (resp. $r < 0$). An easy consequence of FVIII is that the square of every non zero real number is positive. Thus the basic deficiency of real number system is that the equation

$$X^2 + 1 = 0 \tag{1.15}$$

has no solution in real numbers. The irresistible desire to have a number system in which the equation (1.15) has a solution leads us to *expand* the real number system in some way or the other. Such a process is, in a way, not at all new. Recall that, starting with the natural numbers,

$$\mathbb{N} := \{0, 1, 2, \ldots, \}$$

one would construct the negative numbers, so that an equation of the form

$$X + n = 0 \tag{1.16}$$

has a (unique) solution for each natural number n. We merely 'cooked-up' a notation '$-n$' to denote this unique solution of the equation (1.16) and then the usual laws of arithmetic were applied to them. The enlarged set $\mathbb{Z}$ of these numbers were then called *integers*. Similarly, we notice that if m, n are integers, there is no guarantee that the equation

$$nX + m = 0 \tag{1.17}$$

has a solution in integers. Of course, (1.17) makes sense only when $n \neq 0$ and then we demand that (1.17) have solutions. Again, the so called *rational numbers* were created out of integers to satisfy this demand, viz., by introducing certain equivalence classes of ordered pairs of integers (m, n) where $n \neq 0$.[4] Later, a simpler notation was found to denote these equivalence classes, viz., the class of (m, n) is denoted by $\frac{m}{n}$.

We propose to handle the equation (1.15) also in a similar manner, subject to the following **thumb rule:** that whatever *extended system of numbers* we get, the system should satisfy the laws $FI - FVI$. This is indeed essentially the historical way how complex numbers were introduced to humanity. The difference is that we shall put it in the modern language.

Let us make some formal definitions. Recall that a set K together with two binary operations obeying the Rules FI-FVI is called a *field.* Any subset L of K which again obeys the same set of rules on its own, where the two binary operations are taken as in K is called a *sub-field.* Typical example of a field is the set $\mathbb{R}$ of real numbers with the two binary operations as the usual addition and the multiplication. The set $\mathbb{Q}$ of rational numbers then forms a sub-field of $\mathbb{R}$. Observe that the set of irrational numbers does not form a sub-field; neither does the set $\mathbb{Z}$ of integers.

So, we start by introducing a formal symbol $\imath$, as a solution of the equation (1.15), i.e.,

$$\boxed{\imath^2 + 1 = 0; \quad \text{i.e., } \imath^2 = -1.} \tag{1.18}$$

Once we allow the quantity $\imath$ into our number system along with all the real numbers, we are forced to allow quantities such as $a + b\imath$ for all $a, b \in \mathbb{R}$, which are obtained by merely applying the two arithmetic operations of addition and multiplication to all the real numbers together with the newly introduced symbol $\imath$. Since we do not want to expand the system unnecessarily further, let us stop here and check how good is this new collection. Let us denote the set of all these formal expressions of the form $a + b\imath,\ a, b \in \mathbb{R}$ by $\mathbb{C}$. Observe that we are treating a real number a as being equal to the expression $a + 0\imath$. In this way, $\mathbb{R}$ is now a subset of $\mathbb{C}$.

One of the fundamental properties of these formal expressions $a + b\imath$ is that

$$\boxed{a + b\imath = 0 \ \text{ iff } \ a = 0 \ \text{ and } \ b = 0.} \tag{1.19}$$

What should be the quantity $(a + b\imath) + (c + d\imath)$? Since the *thumb rule* should prevail, the first three laws FI, FII and FIII applied appropriately give:

$$\begin{aligned} (a + b\imath) + (c + d\imath) &= a + (b\imath + c) + d\imath = a + (c + b\imath) + d\imath \\ &= (a + c) + (b\imath + d\imath) \\ &= (a + c) + (b + d)\imath. \end{aligned} \tag{1.20}$$

[4] $(m_1, n_1) \simeq (m_2, n_2)$ iff $m_1 n_2 = n_1 m_2$.

Likewise it follows that

$$\begin{aligned}(a+b\imath)(c+d\imath) &= a(c+d\imath)+b\imath(c+d\imath)\\ &= ac+ad\imath+b\imath c+b\imath d\imath\\ &= ac+ad\imath+bc\imath+bd\imath^2\\ &= (ac-bd)+(ad+bc)\imath.\end{aligned} \tag{1.21}$$

The rule (1.19) automatically sets up a bijection[5] between $\mathbb{C}$ and the set of ordered pairs of real numbers

$$a+b\imath \mapsto (a,b).$$

Check that the laws of addition and multiplication, (1.20) and (1.21) correspond to the two operations in (1.1), under this bijection. Thus these algebraic operations are forced on us in a natural way. Even though the law of multiplication (1.21) looks somewhat contorted, it is a logical consequence of th choices we have made.

Now, it is fairly easy to see that these operations of addition and multiplication on $\mathbb{C}$ obey rules FI–FV. One of the simplest consequences of FI is that we can write $a+b\imath$ in different orders such as $a+\imath b$ or $\imath b+a$ etc.. Finally, in order to verify FVI, first observe that $a+b\imath \neq 0$ iff $a^2+b^2 \neq 0$. Thus, for $z=a+b\imath \neq 0$, put

$$w=\frac{a}{a^2+b^2}+\imath\frac{-b}{a^2+b^2}$$

and verify $zw=1$. This means that every non zero complex number has a multiplicative inverse. Thus $\mathbb{C}$ is a field. Clearly, $\mathbb{R}$ is a sub-field of $\mathbb{C}$.

What we have achieved so far may be summed-up in the following:

Theorem 1.2.1 *The set $\mathbb{C}$ of all formal expressions $a+\imath b$ where $a,b\in\mathbb{R}$ forms the smallest field containing $\mathbb{R}$ as a sub-field and in which $\imath$ is a solution of the equation*

$$X^2+1=0.$$

We would like to draw your attention to the word ‘smallest’ in the above theorem: Let K be any field which contains $\mathbb{R}$ as a sub-field and in which equation (1.15) has a solution. It is now only a matter of selecting some notation for such a root and this could as well be $\imath$. Then being a field, K would contain all the elements of the form $a+b\imath$ where a,b are real numbers. This means $\mathbb{C}$ is a sub-field of K. It is precisely in this sense that the word ‘smallest’ is used here.

[5]This was fruitfully noticed by Hamilton which lead him to the discovery of quaternions.

Remark 1.2.1 You may have heard of Cardano[6] in connection with his solution of cubic equations. Perhaps, he is the first one to introduce the complex numbers around 1545, even without realizing it, in proposing and solving the following problem: *Divide 10 into two parts so that the product of these two parts is equal to 40.* As a warming-up exercise, show that this problem has no solution in reals. Cardano strongly believed in this. Instead of taking it as a final answer, Cardano, with great reluctance perhaps, created the complex numbers and then rejected the whole idea by terming it as *fictitious.* His grip over them seems to have been quite shaky, chiefly because he was not psychologically ready to have a number system in which two numbers cannot be compared. Five years later, Bombelli[7] introduced the complex numbers more systematically in his famous book *Algebra*, which he wrote shortly after Cardano's *Ars Magna.* Nevertheless, certain mysticism surrounded the complex numbers. Even Euler, who had great mastery over the complex numbers tried in vain to put an order on the complex numbers. Finally it was the pioneering work of Hamilton [8] which cleared this mystery surrounding the complex numbers and eventually "liberated algebra from arithmetic"[9]

Let us see why Euler was bound to fail in his attempt in putting an order on complex numbers.

If possible, let there be such a total order '$<$' on $\mathbb{C}$ which is compatible with the arithmetic operations, i.e., one which obeys the laws FVII and FVIII. Then it follows that the square of every non zero complex number is positive and hence both ± 1 are positive. That is a contradiction.

Remark 1.2.2 There are many symbols, in the literature, for $\sqrt{-1}$. The most popular symbol is i, which was used by Euler[10], for the first time in 1777. Later Gauss[11] used it systematically. Nevertheless, it took a lot more time to become commonly accepted.

[6] Girolamo Cardano(1501-1576) was a Milanise doctor and a professor of mathematics, known for his scandals, his book *Ars Magna.* He lectured and wrote on mathematics, medicine, astronomy, astrology, alchemy and physics.

[7] Raffael Bombelli was a Bolognese Engineer (1526-1572).

[8] William Rowan Hamilton, an Irish Mathematician (1805-1865), discovered the quaternions in 1843, the first non commutative algebra to be studied. He felt this would revolutionize mathematical physics and he spent the rest of his life working on quaternions.

[9] See Klein: Symmetry?

[10] Leonhard Euler(1707-1783) was a Swiss mathematician, the most productive amongst all the eighteenth century mathematicians. His major contributions are in mechanics, trigonometry, geometry, differential calculus and number theory. His collected works contains 886 books and papers.

[11] Carl Friederick Gauss (1777-1855) was a German, the first modern mathematician. He has been described as the Prince of Mathematics. He worked in a variety of fields both in mathematics and

In electrodynamics, the symbol i is used to denote the strength of the current. That is the reason why the electrical engineers cooked up a different symbol viz. j for $\sqrt{-1}$. However, in aerospace engineering, j stands for $log(-1)$ and hence they started using another symbol $\imath$ (the dot-less i) which was earlier introduced by Dickson[12]. This last symbol has been now adopted by Donald Knuth in Tex and we too have adopted it. Strictly speaking, we should read this symbol *'iota'.* Since this would divert too much from the established practice, we read it as 'eye' to rhyme with 'bye' or 'my'.

Example 1.2.1 Alright. We can now solve (1.15). We can even solve equations of the type

$$z^2 = a + \imath b, \tag{1.22}$$

where $a, b \in \mathbb{R}$, as follows.

We set $z = u + \imath v$. Then (1.22) is equivalent to

$$u^2 - v^2 = a; \quad 2uv = b.$$

Consider the case $b \neq 0$. Then $u \neq 0$. Substituting for v in the first relation from the second, and clearing the denominator, we obtain

$$4u^4 - 4u^2a - b^2 = 0.$$

We can re-write this in the form

$$(2u^2 - a)^2 = a^2 + b^2.$$

Therefore it follows that

$$u^2 - a = \sqrt{a^2 + b^2}.$$

[This is where we assume that every non negative real number has a square root. A rigorous proof of this follows from the intermediate value theorem discussed in theorem 1.5.6 in section 5.] Thus

$$u = \pm\sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}}; v = \pm\sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}}. \tag{1.23}$$

physics. If you have doubts about the originator of some classical result, attribute it to Gauss and there is a good chance that you may be correct.

[12] Leonard E. Dickson, a U. S. mathematician, known for his work on the theory of finite groups.

Here, the sign of v has to be chosen depending on the sign of u and b so that $2uv = b$. Now, check that $z = u + \imath v$, satisfies (1.22). This entire method is known as the *method of completing the squares* and is due to Shridharacharya. [13]

By taking successive square-roots, it follows that we can solve any equation of the form $z^{2^k} = w$. We can also employ this method to solve any quadratic equation with complex coefficients. Write $z^2 + bz + c = (z + b/2)^2 + c - b^2/4$. Now take w so that $w^2 = b^2/4 - c$. Then $z = \pm w - b/2$ are the solutions of $z^2 + bz + c = 0$. All this must be familiar to you from your school days.

Remark 1.2.3 One can go on like this perhaps to some extent, but you will soon perceive that such arithmetic methods are insufficient to solve an arbitrary polynomial equation.[14] What do we do then? Do we go on introducing more and more 'numbers' which are 'solutions' of such equations that we cannot solve? Before doing anything like that we should get hold of a polynomial $p(z)$ which definitely does not have any roots in $\mathbb{C}$.

One of the many wonderful consequences of existence of solution of (1.15) in $\mathbb{C}$ is that

Every non constant polynomial equation in one variable and with coefficients in $\mathbb{C}$ *has a root in* $\mathbb{C}$.

This result is called the Fundamental Theorem of Algebra. Thus there is an enormous gain by the mere introduction of the symbol $\imath$, which perhaps compensates in some way, the loss of total order resulted due to the same. Perhaps, this *'give and take'* aspect is the one that contributes to the entire mysterious beauty of the inter-relationship between real and complex numbers.

Complex analysis offers one of the most elegant proofs of the Fundamental Theorem of Algebra. This and some other proofs will be presented in this book. However, the most elementary proof seems to be the one which uses just a few facts about the real line that you have already learnt and we shall learn this proof in the last section of this chapter.

[13]Sridhara (around 870-930) was an Indian mathematician who wrote on practical applications of algebra and was one of the first to give a formula for solving quadratic equations.

[14]Indeed, Cardano succeeded in solving all cubics by this method. Later Ferrari found a method to solve all quartics i.e., equations of degree four. Collectively these methods are called 'solution by radicals'. Later it was proved that there are general quintics (and therefore polynomials of higher degree) that cannot be solved by radicals (Abel-Ruffini Theorem)

Remark 1.2.4 As explained in remark 1.2.1, the words *real*, *imaginary*, *complex* etc. have to stay for historical reasons only. In fact, the *real numbers* are no more `real` than the *imaginary numbers* and one does not need any more imagination to visualize the *imaginary part* of a complex number than its *real part.* The concept of complex numbers is no more complicated or intriguing than the concept of real numbers. For that matter, even today, all school children are ingrugued by negative numbers. In the first exercise below, we give three illustrations to help resolve any such intrigue with complex numbers.

Exercise 1.2 A mapping $\phi : K \longrightarrow L$ between two fields is called an *isomorphism* if it is a bijection and preserves the two algebraic operations, viz., $\phi(z_1 + z_2) = \phi(z_1) + \phi(z_2)$ and $\phi(z_1 z_2) = \phi(z_1)\phi(z_2)$.

1. Now consider the way a standard book defines complex numbers. On the set $\mathbb{R}^2$ of ordered pairs of real numbers, the operation of addition and multiplication are given by (1.1). Show that these operations make $\mathbb{R}^2$ a field.

2. Let $\mathbb{C}$ denote the field of complex numbers as defined in this section. Consider the bijection

$$(x, y) \mapsto x + \imath y$$

from $\mathbb{R}^2$ to $\mathbb{C}$. Check that the two operations in (1.1) correspond to (1.20), and (1.21) respectively and hence the two fields are isomorphic. Thus, we see that multiplication rule in (1.1) is not a mere concoctment but actually forced on us because of other considerations which are all natural.

   This is the elegant way Hamilton defined complex numbers, bringing down the mystery surrounding the complex numbers to a great extent. On the other hand, this new approach to complex numbers lead him further to discover the quaternions.

3. Another interesting way to obtain the field of complex numbers is as follows: Consider the ring $\mathbb{R}[t]$ of all polynomials in one variable with real coefficients. Consider the ideal $(t^2 + 1)$ and let $K = \mathbb{R}[t]/(t^2 + 1)$ be the quotient ring. Use division algorithm to write any polynomial $p(t)$ in the form

$$p(t) = q(t)(t^2 + 1) + at + b,\ \ a, b \in \mathbb{R}.$$

Use this to show that $p(i) = 0$ iff $t^2 + 1$ divides $p(t)$. Now, consider the function $\sigma : \mathbb{R}[t] \longrightarrow \mathbb{C}$ given

$$p(t) \mapsto p(\imath)$$

i.e., substituting $\imath$ for t in each polynomial. Show that σ defines an isomorphism $\hat{\sigma} : K \longrightarrow \mathbb{C}$.

4. Here is yet another way to see that we do not need to actually construct the complex numbers. Let

$$K_1 = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \; : \; a, b \in \mathbb{R} \right\}$$

be the subset of the *ring* of 2×2 real matrices. Verify that K_1 is closed under matrix addition and multiplication, and thus becomes a subring. Also, show that

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 = - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Finally, show that the mapping

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mapsto a + \imath b$$

defines an isomorphism of K_1 with the field of complex numbers, under which the subset of scalar matrices (obtained by taking $b = 0$ above) is mapped isomorphically onto the field of real numbers.

Thus, we see that there are several ways to 'avoid' construction of complex numbers in an abstract way if that is what we wanted. In other words, $\imath$ is not an *invention* but a *discovery*. **Complex numbers were already there, manifest in so many ways, as ordered pair of reals, as a certain quotient field of the ring of polynomials or as a sub-algebra of 2×2 real matrices, etc..**

1.3 Geometry of Complex Numbers

As we have seen, a complex number $z = x + \imath y$ is uniquely determined by the two real numbers, viz., its real part x and its imaginary part y. Thus we see that the set of complex numbers corresponds naturally with the set of ordered pairs of real numbers

(x, y). These, in turn, correspond to points in the real 2-dimensional Cartesian space $\mathbb{R} \times \mathbb{R}$. Thus a purely real number x corresponds to a point on the $X-$axis (also called the *real axis*). Also a purely imaginary number $\imath y$ corresponds to a point on the $Y-$axis (also called the *imaginary axis*). The number 0 corresponds to (0,0). This Cartesian plane endowed with the structure of complex numbers is called the *Argand*[15] *plane* or the *complex plane.*

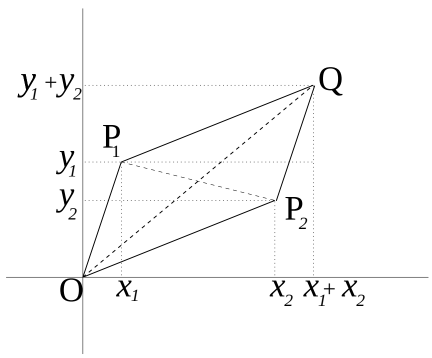


Fig.1

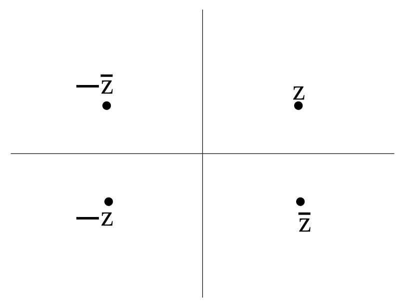


Fig.2

It is easily seen that if z_1 and z_2 are two complex numbers represented by points P_1 and P_2, then $z_1 + z_2$ is represented by the fourth vertex of the parallelogram on the vectors $\overrightarrow{OP}_1$ and $\overrightarrow{OP}_2$.

The 'parallelogram law' that we saw in section 2 is precisely the parallelogram law of the plane geometry viz.,

> The sum of the squares of the lengths of the diagonals
> of a parallelogram is equal to the sum of
> the squares of the lengths of the sides.

Next note that conjugation corresponds to taking reflection about the real axis. In general, the four points $z, \bar{z}, -z, -\bar{z}$ form the four vertices of a rectangle situated symmetrically about the origin.

[15] Jean Robert Argand (1768-1822) of Geneva published an account on graphical representation of complex numbers 1806. However, Casper Wessel (1745-1818) had published the idea of graphical representation of complex numbers in a Danish journal, in 1798, which went unnoticed. Also, Gauss himself had already used such representation of complex numbers in his thesis (1798) in which he gave the first correct proof of Fundamental Theorem of Algebra. Thus, it would be appropriate to use the expression *Wessel-Gauss-Argand diagram* instead.

The geometric representation of complex numbers has many advantages both for the study of the complex numbers as well as that of the (plane) geometry, illustrated with a sequence of exercises at the end of this section. To begin with, we give another representation of complex numbers.

Polar Coordinates: Once we fix the origin 0 and the two coordinate axes, every point z in the plane other than the origin is at a positive distance r from 0. Moreover, the line segment $[0, z]$ makes a certain angle θ with the positive real axis. The pair (r, θ) is called the polar representation of the point $z \neq 0$.[16]

For a point $z = (x, y)$ we know from trigonometry that

$$\boxed{x = r\cos\theta; \qquad y = r\sin\theta.} \tag{1.24}$$

Let us temporarily set-up the notation

$$\boxed{E(\theta) := \cos\theta + \imath\sin\theta.} \tag{1.25}$$

Then the complex number $z = x + \imath y$ takes the form

$$z = r(\cos\theta + \imath\sin\theta) = rE(\theta).$$

Of course, here we have the real number $r = |z| \geq 0$, and θ is defined only for $r > 0$. We call the angle θ made by the line segment $[0, z]$ with the real axis, the *argument* or *amplitude* of z and write $\theta = \arg\ z$. This angle is measured in the counter clock-wise direction.

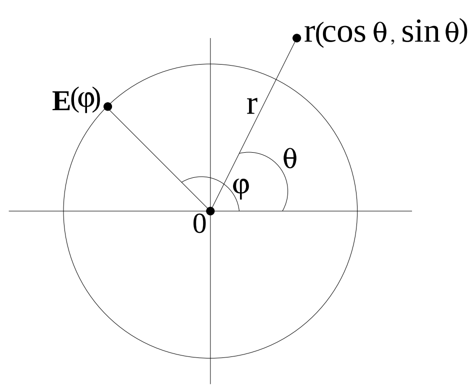


Fig. 3

[16]Here is another point that intrigued me as a strudent. More on it later.

Now let $z_1 = r_1E(\theta_1), \ \ z_2 = r_2E(\theta_2)$. Using additive identities for sine and cosine viz.,

$$\boxed{\boxed{\begin{array}{rcl} \sin(\theta_1+\theta_2) & = & \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2, \\ \cos(\theta_1+\theta_2) & = & \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2, \end{array}}} \tag{1.26}$$

we obtain

$$\boxed{\boxed{z_1z_2 = r_1r_2E(\theta_1+\theta_2).}} \tag{1.27}$$

If we further remind ourselves that the argument can take values (in radians) between 0 and 2π, then the above identity tells us that $\arg(z_1z_2) = \arg z_1 + \arg z_2 \pmod{2\pi}$ provided $z_1 \neq 0, \ \ z_2 \neq 0$.

Now, we can justify the name *cosine rule* for (B4) in section 1.2. Let $z_j = r_jE(\theta_j)$ for $j = 1, 2$, be two non zero complex numbers and let θ be the angle between the vectors represented by them. Then $z_1\bar{z}_2 = r_1r_2E(\theta_1 - \theta_2)$ and hence $\Re(z_1\bar{z}_2) = r_1r_2\cos\theta$. Thus,

$$\boxed{\boxed{\cos\theta = \frac{\Re(z_1\bar{z}_2)}{|z_1z_2|}.}} \tag{1.28}$$

Now, we can rewrite the cosine rule as:

$$|z_1+z_2|^2 = r_1^2 + r_2^2 + 2r_1r_2\cos\theta. \tag{1.29}$$

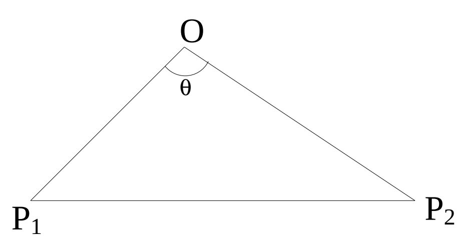


Fig. 4

In the Fig. 4, the angle between OP_1 and OP_2 is equal to θ. The square of the length of P_1P_2 is given by

$$|z_1-z_2|^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\theta. \tag{1.30}$$

This is precisely the cosine rule that you have come across in your school geometry. (Given a triangle $\Delta(ABC)$, choose the vertex A as the origin, $B = z_1$ and $C = z_2$.)

Note that by putting $\theta = \pi/2$ in (1.29), we get Pythagoras theorem.

Remark 1.3.1 We have deliberately chosen to depend upon our geometric intuition in the treatment of the concept of argument. It is possible to give a rigorous analytic treatment in several ways. The reader may refer to chapter 1 section 3 of [Car] for an elegant account of this OR she may choose to work through the exercise 15 in Miscellaneous Exercises to Ch. 2.

Observe that given $z \neq 0$, $\arg z$ is a multi-valued function. Indeed, if θ is one such value then all other values are given by $\theta + 2\pi n$, where $n \in \mathbb{Z}$. Thus to be precise, we have

$$\arg\ z = \{\theta + 2\pi n\ :\ n \in \mathbb{Z}\}$$

(In Greek mythology 'Argus' is a 100-eyed giant!) This is the first natural example of a ' multi-valued function'. We shall come across many multi-valued functions in complex analysis, all due to this nature of $\arg z$. However, while carrying out arithmetic operations we must 'select' a suitable value for arg from this set. One of these values of $\arg z$ which satisfies $-\pi < \arg z \leq \pi$ is singled out and is called the *principal value* of $\arg\ z$ and is denoted by $\mathrm{Arg}\ z$. Thus

$$\boxed{\boxed{-\pi < \mathrm{Arg}\ z \leq \pi.}} \tag{1.31}$$

We would like to draw your attention to the unfortunate fact that there is no agreement, in literature, on the use of the notations $\arg z$ and $\mathrm{Arg}\, z$. They are used interchangeably and moreover, there is no agreement even on the choice of the principal value –a strong contender for the choice that we have made being $0 \leq \theta < 2\pi$. So, it is all the more important for you to realize that the freedom for such a choice can actually be useful and by being little careful, we can avoid pitfalls which are consequences of this ambiguity.

Remark 1.3.2 *Roots of complex numbers:* Thanks to our geometric understanding, we can now show that the equation

$$X^n = z \tag{1.32}$$

has exactly n roots in $\mathbb{C}$ for every non zero $z \in \mathbb{C}$. Suppose w is a complex number that satisfies the equation (in place of X,) we merely write

$$z = rE(\mathrm{Arg}\ z), \quad w = sE(\mathrm{Arg}\, w).$$

Then we have,

$$s^n E(n\mathrm{Arg}\, w) = w^n = z = rE(\mathrm{Arg}\, z)$$

Therefore we must have $s = \sqrt[n]{r} = \sqrt[n]{|z|}$ and $\arg w$ will contain the values

$$\frac{\mathrm{Arg}\, z}{n} + \frac{2k\pi}{n}, \quad k = 0, 1, \ldots, n-1.$$

Thus we see that (1.32) has n distinct solutions. The solution $\sqrt[n]{|z|}E(\frac{\mathrm{Arg}\, z}{n})$ is called the *principal value* of the n^{th} root function and is denoted by $\sqrt[n]{z}$.

Remark 1.3.3 de Moivre's[17] Law Now observe that, by putting $r_1 = r_2 = 1$ in (1.27) we obtain:

$$E(\theta_1 + \theta_2) = E(\theta_1)E(\theta_2).$$

Putting $\theta_1 = \theta_2 = \theta$ and applying the above result repeatedly, we obtain

$$E(n\theta) = E(\theta)^n.$$

This is restated in the following:

$$\boxed{\texttt{de Moivre's Law:}\ \cos n\theta + \imath \sin n\theta = (\cos\theta + \imath\sin\theta)^n.} \tag{1.33}$$

Example 1.3.1 For any n the principal value of the n^{th}-root function for $z = 1$ is 1. The three cube roots of unity are

$$1, \cos\frac{2\pi}{3} + \imath\sin\frac{2\pi}{3}, \cos\frac{4\pi}{3} + \imath\sin\frac{4\pi}{3}$$

which we can simplify as:

$$1, \frac{-1 + \imath\sqrt{3}}{2}, \frac{-1 - \imath\sqrt{3}}{2}.$$

Remark 1.3.4 Geometric Multiplication of Complex Numbers: Starting with $P = z, Q = w$ three pairs of similar triangles are constructed as shown in the figure below:

$$0QR\ |||\ 01P, \quad 0S1\ |||\ 01Q \quad 0ST\ |||\ 01P.$$

The points R, S and T represent zw, w^{-1}, zw^{-1} respectively. We leave it you to figure out why.

[17] Abraham de Moivre(1667-1754) was a French mathematician. He also worked in Probability theory.

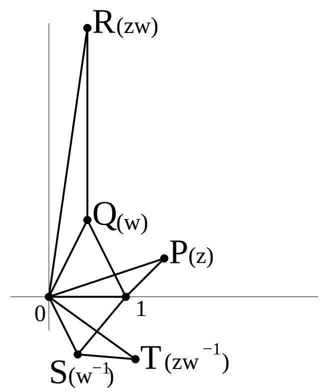


Fig. 5 Geometric way of multiplication

Observe that given $0 \neq \lambda \in \mathbb{C}$, the assignment $z \mapsto \lambda z$ defines a linear map $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$, which is a composite of a *rotation* (through an angle $\theta = \operatorname{Arg} \lambda$) and a *dilation* or a *scaling* (by a factor $r = |\lambda|$). Such linear maps $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$ are called *similarities.*

The dot product Think of two complex numbers $z_j = (x_j, y_j), j = 1, 2$ as vectors in $\mathbb{R}^2$. Then their dot product is given by $x_1x_2 + y_1y_2$. This can be rewritten in terms of multiplication of complex numbers as:

$$z_1 \cdot z_2 = \Re(z_1\bar{z}_2) = \Re(\bar{z}_1z_2). \tag{1.34}$$

Similarly, verify that the cross product can be rewritten in the form

$$z_1 \times z_2 = \Im(\bar{z}_1z_2)\mathbf{k} \tag{1.35}$$

where $\mathbf{k}$ is the unit vector $(0, 0, 1)$ in $\mathbb{R}^3$. Thus the geometric meaning of $\Re(z_1\bar{z}_2) = 0$ is that the vectors z_1 and z_2 are perpendicular to each other. Similarly, the geometric meaning of the condition $\Im(z_1\bar{z}_2) = 0$ is that the complex numbers z_1, z_2 are real multiples of each other.

Equations of lines and circles Let $ax + by + c = 0$ represent a line in Cartesian coordinates. Without loss of generality we may assume that $a^2 + b^2 = 1$. Put $w = a + \imath b; z = x + \imath y$. Then $ax + by = \Re(\bar{w}z) = (w\bar{z} + \bar{w}z)/2$. Thus, we see that the general equation of a line in the plane is of the form

$$\boxed{w\bar{z} + \bar{w}z = 2t, \quad |w| = 1, \quad t \in \mathbb{R}.} \tag{1.36}$$

Combined with (1.34), it follows that the line represented by (1.36) is perpendicular to the vector w. The real number t on the rhs of (1.36) represents 'how far' the line is from the origin. Also, we can now write down the equation of a line which passes through two given w_1, w_2 as follows: This line will be perpendicular to $w = \imath(w_1 - w_2)$. The value of t can be computed by using the fact that w_1 (or w_2) lies on the line. Upon simplification, this gives

$$\Im(\bar{w}_1 z) - \Im(\bar{w}_2 z) = \Im(\bar{w}_1 w_2). \tag{1.37}$$

The equation of a circle is easier : if the center is c and the radius is r, we have,

$$\boxed{\boxed{(z-c)(\overline{z-c}) = r^2, \quad z \in \mathbb{C}, r > 0.}} \tag{1.38}$$

Equivalently

$$\boxed{\boxed{|z-c| = r, \ \ z \in \mathbb{C}, r > 0.}} \tag{1.39}$$

Definition 1.3.1 By a **rigid motion** (RM) or an **isometry** of the plane, we mean a mapping $f : \mathbb{C} \to \mathbb{C}$ which preserves distances, i.e., $|f(z) - f(w)| = |z - w|$ for all $z, w \in \mathbb{C}$.

Remark 1.3.5 Identity, translation, rotation about a point and reflection in a line are obvious examples of rigid motions. The first three preserve the sense of orientation of the plane and the third one changes it. The first one keeps all points fixed. The second one has no fixed point but it preserves orientation. The third one fixes exactly one point and also preserves the orientation. The fourth one fixes exactly a line and (hence) changes the orientation. Are there other rigid motions such as those which fix no points and change the orientation? Of course, a composite of finitely many rigid motions is again a rigid motion and so we can take composites as many of the above cited examples. We now have a number of questions: *Is every rigid motion a composite of these ones? Do these composites have any strange geometric properties that we have not seen in the above examples? It is easy to check that every RM is a 1-1 mapping. Is it onto? etc.* We shall investigate these questions right now. Let us first take a closer look at some of these examples.

Example 1.3.2 Rotation about a point w by an angle θ. We have observed that multiplication by $E(\theta)$ results in rotating vectors (originating at 0) through an angle θ. Thus, to understand the effect of rotation about a point w, we look at the vector $z - w$

placed at 0. This will be rotated through an angle θ which yields the vector $E(\theta)(z-w)$. This has to be placed back at the point w and that gives us the point $E(\theta)(z-w)+w$. Thus, the rotation about w through an angle θ is given by

$$\boxed{z \mapsto E(\theta)(z-w)+w = E(\theta)z + w(1-E(\theta))} \tag{1.40}$$

which is a rotation followed by a translation. This can fruitfully be employed to find the point around which the rotation is being performed by a RM of the form $z \mapsto E(\theta)z+b$. viz., $w = (1-E(\theta))^{-1}b$. From this it also follows that composite of two rotations is a rotation and composite of a rotation and a translation in whichever order is again a rotation.

At this stage, it is appropriate that we make it clear what we mean by orientation. On the real line, the notion of orientation coincides with the notion of directions, positive or negative. In the case of the plane, suffices it to say that this just refers to the choice that we have made in measuring the angle out of two possible ways, viz., clockwise or anti-clockwise.

Example 1.3.3 Reflection in a Line: Let

$$w\bar{z} + \bar{w}z = 2t$$

represent a line. This line is perpendicular to w. Therefore, if $R(z) := z^*$ denotes the image of a variable point z under the reflection in L, then the line segment $[z^*, z]$ is parallel to w and is bisected by L, i.e, $\frac{z^*+z}{2}$ lies on the line L. Therefore, we obtain,

$$z^* - z = sw, s \in \mathbb{R}; \quad \& \quad w(\overline{z^*+z}) + \bar{w}(z^*+z) = 4t.$$

Substitute $z^* = z + sw$ in the latter and use the fact $w\bar{w} = 1$ to obtain $s = 2t - (w\bar{z}+\bar{w}z)$. Simplify to get

$$\boxed{z^* = 2wt - w^2\bar{z}.} \tag{1.41}$$

Equivalently, by multiplying by $\bar{w}$, we get

$$\boxed{w\bar{z} + \bar{w}z^* = 2t.} \tag{1.42}$$

Definition 1.3.2 By a *glide reflection* we mean a RM which is a reflection in a line L followed by a translation by a **non zero** vector parallel to L.

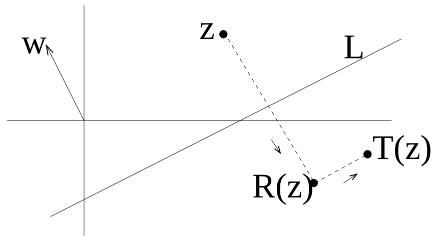


Fig. 6 Reflection and glide reflection

It is easy to see that a glide-reflection does not have any fixed points and does not preserve the orientation. The converse follows from what we see below. In the Fig. 6, the line L is perpendicular to the unit vector w; R and T represent the reflection and a glide-reflection.

Theorem 1.3.1 *Let $f : \mathbb{C} \to \mathbb{C}$ be a rigid motion. Then there exist unique $a, b \in \mathbb{C}$ with $|a| = 1$ such that*

$$f(z) = az + b, \forall z \in \mathbb{C} \ OR \ f(z) = a\bar{z} + b, \forall z \in \mathbb{C}.$$

Proof: Put $b = f(0)$ and define $g(z) = f(z) - b$. Then g is also a RM and $g(0) = 0$. Now $|g(1)| = 1$. So, put $a = g(1)$ and define $h(z) = a^{-1}g(z)$. Then h is a RM and $h(0) = 0$ and $h(1) = 1$. Therefore $h(\imath) = \pm\imath$.
Case 1: Assume $h(\imath) = \imath$. Now consider any $z = x + \imath y \in \mathbb{C}$ and put $h(z) = u + \imath v$. Then it follows that
$u^2 + v^2 = x^2 + y^2; \quad (u-1)^2 + v^2 = (x-1)^2 + y^2; \quad u^2 + (v-1)^2 = x^2 + (y-1)^2.$
Solving these, yields, $u = x, v = y$. Thus $h(z) = z, \forall z \in \mathbb{C}$. This is the same as saying $f(z) = az + b, \forall z \in \mathbb{C}$.
Case 2: Assume that $h(\imath) = -\imath$. Put $\bar{h}(z) = \overline{h(z)}$. Then $\bar{h}$ is a RM and $\bar{h}(0) = 0, \bar{h}(1) = 1, \bar{h}(\imath) = \imath$. So, we are in case 1. ♠.

Throughout this book, we shall use the symbol ♠ to indicate the end a proof, as done above.

Theorem 1.3.2 *Let $f : \mathbb{C} \to \mathbb{C}$ be a rigid motion.*
(i) *Suppose f fixes two distinct points. Then all points on the line passing through these two points are also fixed by f.*
(ii) *Suppose f fixes three non collinear points. Then $f = Id$.*
(iii) *Suppose f fixes an entire line L. Then it is either Id or the reflection in that line.*
(iv) *Suppose f fixes exactly one point. Then it is a rotation around that point.*
(v) *Suppose f fixes no points. Then either f is a translation or a glide reflection.*

Proof: We make use of the above theorem here.
(i) If z_1, z_2 are such that $az_i + b = z_i, i = 1, 2$ then for any real numbers t_1, t_2 such that $t_1 + t_2 = 1$, we have

$$a(t_1z_1 + t_2z_2) + b = t_1(az_1 + b) + t_2(az_2 + b) = t_1z_1 + t_2z_2.$$

This takes care of the case when $f(z) = az + b$. The case when $f(z) = a\bar{z} + b$ is similar.
(ii) Argue as in (i).
(iii) We may choose L to be the real axis. And then depending on whether $f(\imath) = \pm\imath$, it follows that $f(z) = z$ or $f(z) = \bar{z}$.
(iv) Choose the fixed point of f as the origin. Then $f(0) = 0$. This means $b = 0$. Suppose f is of the second form $f(z) = a\bar{z}$, with $a = E(\theta)$. Then $f(rE(\theta/2)) = rE(\theta/2)$, This means that f fixes a whole line. This is ruled out. Therefore f is of the form: $f(z) = E(\theta)z$ for some θ.
(v) First consider the case $f(z) = az + b$. If $a \neq 1$ then we can solve for $f(z) = z$, viz, $z = \dfrac{b}{1-a}$. Therefore $a = 1$ and f is a translation.

Now consider the case, $f(z) = a\bar{z} + b$. Put $a = -w^2$ for some unit vector w. Write $b = tw + s\imath w,\ t, s \in \mathbb{R}$. Then

$$f(z) = (-w^2\bar{z} + tw) + s\imath w.$$

From (1.41), it follows that f is a reflection in a line perpendicular to w followed by a translation by $s\imath w$ which is parallel to the line of reflection. It also follows that $s \neq 0$, for otherwise, f would be just a reflection and will fix a whole line. ♠

Remark 1.3.6 Thus we see that there are just five different classes of rigid motions of the plane: Identity, translation, rotation, reflection, glide reflection.

We shall end this section with a story due to George Gamow, the well-known physicist and an ingenious story-teller. We quote from his book:

ONE TWO THREE $\cdots$ **INFINITY** (pp. 44-45).

There was a young and adventurous man who found among his great-grand father's papers a piece of parchment that revealed the location of a hidden treasure. The instructions read:

" Sail to $\cdots$ North latitude and $\cdots$ West longitude where thou wilt find a deserted island. There lieth a large medow, not pent, on the north shore of the island where standeth a lonely oak and a lonely pine. There thou wilt see also an old gallows on

which we once were wont to hang traitors. Start thou from the gallows and walk to the oak counting thy steps. At the oak thou must turn right *by a right angle and take the same number of steps. Put here a spike in the ground. Now must thou return to the gallows and walk to the pine counting thy steps. At the pine thou must turn* left *by a right angle and see that thou takest the same number of steps, and put another spike into the ground. Dig half-way between the spikes; the treasure is there."*

The story is that when the young man finally landed on the island, even though he could find the medow and the two trees as described, the old gallows had totally disappeared! The adventurous man fell into a despair and started digging at random here and there in the vast island and finally had to return empty handed.

Now, the point Gamow wants to make is that if the young man knew a bit of mathematics, particularly the use of complex numbers, he could have found the treasure. We only add that, if the man had not despaired, just made a single attempt to guess the location of the gallows and carried out the instructions given in the parchment, he would have got the treasure as well as the great satisfaction (perhaps falsely) of guessing the position of the gallows correctly. Try to figure it out yourself before reading the solution so that you will have the satisfaction of finding the treasure. We would like to emphasis the fact that, it is not very difficult to solve this problem through elementary school geometry, either. So, you see that it is unlikely that the young man in the story did not know even that much mathematics and still could navigate to the island. Of course, if you despair, no amount of mathematics will help you!

[Solution: Let us represent the map of the island by complex numbers. Of course, we are free to choose our axes and what is better than to choose the line joining the two trees as the real axis! Now, clearly, half-way between the two trees should be a good choice for the origin. Then it really should not matter whether the oak or the pine is chosen as the point 1 say, the pine. Then naturally the oak will refer to -1. Let now Γ denote the position of the gallows, which none of us know. The point is that it does not matter: carry out the rest of the instructions and you arrive at an answer independent of this unknown quantity. We feel that you should still try this problem on your own. At this stage we shall give you a hint: use the fact that multiplication by $\imath$ corresponds to turning a vector around a right angle in the anti-clockwise direction. Read further, only after you have tried enough.

The position S_1 of the first spike is found as follows: The vector representing the distance and the direction from the gallows to the oak is $-1 - \Gamma$. Therefore, the vector representing the direction and the distance from the oak to the first spike is got by multiplying by $-\imath$, viz., $\imath(1+\Gamma)$. Since, this vector has to originate at the oak, we see that $S_1 = \imath(1+\Gamma)-1$. Likewise, the position of the second spike is given by $S_2 = \imath(1-\Gamma)+1$. The midpoint of the line segment $[S_1, S_2]$ is then $(S_1+S_2)/2 = \imath$. And **that is where the treasure is!**

The solution of this problem by method of school geometry is completely left to you. See the solution set for some hint. For those of you who would to read more about plane geometry vis-a-vis complex numbers, I recommend [Sh-2].

Exercise 1.3

1. Express the following complex numbers in the polar form.
   (a) $1+\sqrt{3}\imath$; (b) $\frac{1}{2}+\frac{\imath}{2}$; (c) $\imath\sqrt{2}$; (d) -2;

2. Represent the following subsets of the plane by shading the region.
   (a) $\Re(z) \geq 2$; (b) $\Re(z^2) \leq \alpha$; (c) $\Im(z^2) \leq \alpha$;
   (d) $|z^2-2| \leq 1$; (e) $\left|\frac{1}{z}\right| < 1$; (f) $\left|\frac{z-1}{z+1}\right| \leq 1$.

3. Plot all the values of the indicated roots in each case.
   (a) $\sqrt[3]{1}$; (b) $\sqrt[3]{\imath}$; (c) $\sqrt[4]{-16}$.

4. Put $z_j = R_j E(\theta_j), j = 1, 2$, for $R_1 > R_2 > 0$. Show that
$$\Re\left(\frac{z_1+z_2}{z_1-z_2}\right) = \frac{R_1^2-R_2^2}{R_1^2+R_2^2-2R_1R_2\cos(\theta_1-\theta_2)}.$$

5. Let z_1, z_2 be any two non zero complex numbers and let $\tau = \cos\theta$ where θ is the angle between the two vectors represented by z_1, z_2. Show that
$$|z_1+z_2| \geq \sqrt{\left(\frac{1+\tau}{2}\right)}(|z_1|+|z_2|).$$

6. Solve the following equations for z :
   (a) $|z|-z = 1+\imath$; (b) $|z|+z = 1+\imath$. (c) $z-|z| = 1+\imath$.

7. We say z is an n^{th} root of unity if $z^n = 1$. Show that
   (i) If z is a root of unity then $|z| = 1$.
   (ii) If z and w are roots of unity then so is zw^{-1}. Show that the set of all n^{th} roots of unity is the vertex set of a regular n-gon.

8. Find all the fifth roots of $\imath$ and write down the values of their Arg.

9. If $1 \neq \omega \in \mathbb{C}$ is such that $\omega^3 = 1$, show that $1 + \omega + \omega^2 = 0$. More generally, if $z^n = 1, \quad z \neq 1$, then show that $1 + z + \cdots + z^{n-1} = 0$.

10. In the picture below, the following pairs of triangles are similar:

$$\Delta 01P|||\Delta 0QR, \Delta 01S|||\Delta 0Q1, \Delta 01P|||\Delta 0ST, \Delta PUQ|||\Delta Q0P, \Delta 01S|||\Delta SV1.$$

If $P = z, Q = w$ what are R, S, T, U, and V?

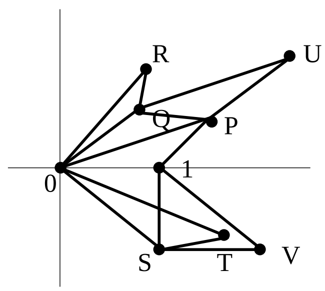


Fig. 7 Geometric way of doing Arithmetic

11. Given a regular $n-$gon with its center at the origin and one of its vertex at the point z find the locations for other vertices.

12. Use geometric arguments to show that
(i) $\left|\dfrac{z}{|z|} - 1\right| \leq |\arg z|, \quad z \neq 0.$
(ii) $|z - 1| \leq |\,|z| - 1| + |z||\arg z|.$
(iii) $|z_1 + z_2| = \left|\dfrac{|z_1|}{|z_2|}z_2 + \dfrac{|z_2|}{|z_1|}z_1\right|, \quad z_1, z_2 \neq 0.$

13. Given a regular n-gon with z_1 and z_2 as adjacent vertices, find all possible locations for
(a) the center c of the polygon;
(b) the vertex w which is adjacent to z_2 and different from z_1.

14. Show that a rotation followed by a translation or vice versa produces a rotation.

15. Show that a reflection followed by a translation or vise versa produces a reflection or a glide-reflection.

16. Show that a reflection followed by a rotation or vice versa produces a glide-reflection in general. When does this produce a reflection?

17. We have derived theorem 1.3.1 using theorem 1.3.3. Prove theorem 1.3.1 directly by geometric methods and then derive theorem 1.3.3 from it.

1.4 Sequences and Series

This is a brief summary of the theory of convergence of sequences and series. We assume that you are already familiar with the general theory of convergence of sequences and series, such as elementary properties, various convergence tests etc.. So, here we recall them very briefly to the extent that is needed for immediate purpose. Some other results are summarized in the form of easily doable exercises at the end of the section. There are several good books from which you can learn this topic better. One such reference is [R]. For the time being, going through the material here should be enough.

We shall only deal with sequences and series of complex numbers, even though most of these results about sequences hold in any metric space, with minor and obvious modifications.

Let $\mathbb{N}$ denote the set of natural numbers,

$$\mathbb{N} := \{1, 2, 3, \ldots\}.$$

Definition 1.4.1 By a *sequence* in a set A, we mean a mapping $f : \mathbb{N} \longrightarrow A$. It is customary to denote a sequence f by $\{s_n\}$ where, $s_n := f(n)$. A sequence $\{z_n\}$ of numbers is said to be *convergent* to the *limit* w if for every $\epsilon > 0$, there exists an integer n_0 such that for all $n \geq n_0$, we have,

$$|z_n - w| < \epsilon.$$

Remark 1.4.1

1. It follows that the limit of a sequence, if it exists, is unique. For, if w_1 and w_2 are two limits of a sequence $\{z_n\}$ then, given $\epsilon > 0$ we can choose n_0 as above, so that for $n \geq n_0$ we have, $|z_n - w_1| < \epsilon$ and $|z_n - w_2| < \epsilon$. Hence, $|w_1 - w_2| < 2\epsilon$. Since $\epsilon > 0$ is arbitrary, we must have $w_1 = w_2$.

2. We use the following two notations

$$\lim_{n \longrightarrow \infty} z_n := w, \qquad \text{OR} \qquad z_n \longrightarrow w,$$

to represent the limit of the sequence $\{z_n\}$ or to indicate that the sequence converges to w.

3. If a sequence is not convergent then it is said to be *divergent.*

4. Amongst divergent sequences of real numbers $\{s_n\}$, there is an important subclass. We say $s_n \to +\infty$ (respectively, $-\infty$) if given any positive M there exists n_0 such that $s_n \geq M$ (respectively, $s_n \leq -M$) for all $n \geq n_0$. We may then say that s_n convergres to $+\infty$ or $-\infty$ accordingly. Only those divergent sequences which are not even 'convergent' to $\pm\infty$ are called *oscillatory* sequences.

5. However, for a sequence of complex numbers $\{z_n\}$, we say $z_n \to \infty$ iff $|z_n| \to +\infty$.

6. If $\{a_n\}, \{b_n\}$ are convergent sequences and $z \in \mathbb{C}$, then the sequences $\{a_n + b_n\}$ and $\{za_n\}$ are also so with limits given by

$$\lim_{n\longrightarrow\infty}(a_n + b_n) = \lim_{n\longrightarrow\infty} a_n + \lim_{n\longrightarrow\infty} b_n; \quad \lim_{n\longrightarrow\infty} za_n = z \lim_{n\longrightarrow\infty} a_n.$$

7. It is a direct consequence of Exercise 1.1.3 that if $z_n = a_n + \imath b_n$, where $a_n, b_n \in \mathbb{R}$, then $z_n \longrightarrow w$ iff $a_n \longrightarrow \Re(w)$ and $b_n \longrightarrow \Im(w)$.

8. The following criterion for convergence of sequences is one of the immediate consequences of the construction of real numbers.

Theorem 1.4.1 *A sequence $\{z_n\}$ of numbers is convergent iff for every $\epsilon > 0$ there exists n_0 such that for all $n, m \geq n_0$, we have $|z_n - z_m| < \epsilon$.*

Remark 1.4.2 In an arbitrary metric space also, we have similar definitions, where the modulus is replaced by the distance function. A sequence satisfying the condition in the above theorem is called a *Cauchy sequence.* The theorem can then be stated as: *a sequence of real or complex numbers is Cauchy iff it is convergent.*

As an important application of sequences, we have:

Theorem 1.4.2 *Let $f : X \longrightarrow \mathbb{C}$ be any function defined on a subset of $\mathbb{C}$. For any $z \in X$, f is continuous at z iff for every sequence $\{z_n\}$ in X such that $z_n \to z$, we have, $f(z_n) \to f(z)$.*

Of course, you may please read definition 1.5.1 for the definition of a continuous function. We need to recall another elementary fact about sequence of real numbers.

Definition 1.4.2 Given a sequence $\{b_n\}$ of real numbers the limsup of the sequence is defined by

$$\limsup_n \{b_n\} := \lim_n (sup\{b_n, b_{n+1}, \ldots\}). \tag{1.43}$$

Remark 1.4.3 Note that in (1.43), we have to allow sequences to take values in $[-\infty, \infty]$, even if we are only interested in real valued sequences. Likewise, here we allow $\pm\infty$ to be genuine limits.

Two important properties of the limsup are the following:
(**Limsup-I**) If $\alpha > \limsup_n \{b_n\}$ then there exists n_0 such that for all $n \geq n_0$ $b_n < \alpha$.
(**Limsup-II**) If $\beta < \limsup_n \{b_n\}$ then there exists infinitely many n_j such that $b_{n_j} > \beta$.
Indeed, lim sup can be characterized by these two properties.

It is important to note that lim sup always exists and can be any value in $[-\infty, \infty]$. When a sequence is convergent (including $\pm\infty$) the $\limsup_n$ of the sequence will be equal to the limit.

This is also the same as the least upper bound of the set of limits of all convergent subsequences of $\{b_n\}$.

Exactly similar way lim inf is also defined and it has similar properties.

We shall leave the proof of the following theorem as an exercise to you.

Theorem 1.4.3 *A sequence $\{b_n\}$ of real numbers is non-oscillatory iff*

$$\limsup_n \{b_n\} = \liminf_n \{b_n\}$$

and in that case, this common value is equal to $\lim_n b_n$.

Remark 1.4.4 Given two numbers, we can add them to get another number. Repeatedly carrying out this operation allows us to talk about sums of any finitely many numbers. We would like to talk about the 'sum' of infinitely many numbers as well. A natural way to do this is to label the given numbers, take sums of first n of them and look at the 'limit' of the sequence of numbers so obtained.

Thus given a (countable) collection of numbers, the first step is to label them to get a sequence $\{s_n\}$. In the second step, we form another sequence: the sequence of partial sums $t_n = \sum_{k=1}^n s_k$. Observe that the first sequence $\{s_n\}$ can be recovered completely from the second one $\{t_n\}$. The third step is to assign a limit to the second sequence provided the limit exists. This entire process is coined under a single term 'series'. However, below, we shall stick to the popular definition of a series.

Definition 1.4.3 By a *series* of real or complex numbers we mean a formal infinite sum:

$$\sum_n s_n := s_0 + s_1 + \cdots + s_n + \cdots$$

Of course, it is possible that there are only finitely many non zero terms here. The *sequence of partial sums* associated to the above series is defined to be $t_n := \sum_{k=1}^{n} z_k$. We say the series $\sum_n s_n$ is *convergent* to the sum s if the associated sequence $\{t_n\}$ of partial sums is convergent to s. In that case, we say s is the *sum* of the series and write

$$\sum_n z_n := s.$$

Once again it is immediate that if $\sum_n z_n$ and $\sum_n w_n$ are convergent series then for any complex number λ, we have, $\sum_n \lambda z_n$ and $\sum_n (z_n + w_n)$ are convergent and

$$\boxed{\boxed{\sum_n \lambda z_n = \lambda \sum_n z_n; \quad \sum_n (z_n + w_n) = \sum_n z_n + \sum_n w_n.}} \tag{1.44}$$

Cauchy's convergence criterion can be applied to series also. This yields:

Theorem 1.4.4 *A series $\sum_n z_n$ of real or complex numbers is convergent iff for every $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0$ and for all $p \geq 0$, we have,*

$$|z_n + z_{n+1} + \cdots + z_{n+p}| < \epsilon.$$

Indeed, all notions and results that we have for sequences have corresponding notions and results for series also, via the sequence of partial sums of the series. Thus, once a result is established for a sequence, the corresponding result is available for series as well and vice versa, without specifically mentioning it.

It follows that if a series is convergent, then its n^{th} term z_n tends to 0. However, this is not a sufficient condition for convergence of the series, as illustrated by the series $\sum_n \frac{1}{n}$.

It is fairly obvious that if $0 \leq a_n \leq b_n$ and $\sum_n b_n$ is convergent then so is $\sum_n a_n$. This result is called comparison test and is so fundamental that we often forget to mention it. Another important convergence test is the so called

Ratio Test: If $\{a_n\}$ is a sequence of positive terms such that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = r < 1,$$

then $\sum_n a_n$ is convergent.

To see this, choose s so that $r < s < 1$. Then there exists n_0 such that $\frac{a_{n+1}}{a_n} < s$ for all $n \geq n_0$. This implies $a_{n_0+k} < a_{n_0} s^k, k \geq 1$. Since the geometric series $\sum_k s^k$ is convergent, the convergence of $\sum_n a_n$ follows.

Definition 1.4.4 A series $\sum_n z_n$ is said to be absolutely convergent if the series $\sum_n |z_n|$ is convergent.

Again, it is easily seen that an absolutely convergent series is convergent, whereas the converse is not true as seen with the standard example $\displaystyle\sum_n (-1)^n \frac{1}{n}$. The notion of absolute convergence plays a very important role throughout the study of convergence of series. As an illustration we shall obtain the following useful result about the convergence of the product series.

Definition 1.4.5 Given two series $\sum_n a_n, \sum_n b_n$, the Cauchy product of these two series is defined to be $\sum_n c_n$, where $c_n = \sum_{k=0}^n a_k b_{n-k}$.

Theorem 1.4.5 *If $\sum_n a_n, \sum_n b_n$ are two absolutely convergent series then their Cauchy product series is absolutely convergent and its sum is equal to the product of the sums of the two series:*

$$\sum_n c_n = \left(\sum_n a_n\right)\left(\sum_n b_n\right). \tag{1.45}$$

Proof: We begin with the remark that if both the series consist of only non negative real numbers, then the assertion of the theorem is obvious. We shall use this in what follows.

Consider the remainder after $n-1$ terms of the corresponding absolute series:

$$R_n = \sum_{k \geq n} |a_k|; \quad T_n = \sum_{k \geq n} |b_k|.$$

Clearly,

$$\sum_{n \geq 0} |c_n| \leq \sum_{k \geq 0} \sum_{l \geq 0} |a_k||b_l| = R_0 T_0.$$

Therefore the series $\sum_n c_n$ is absolutely convergent. Inorder to compute $\sum_n c_n$, consider,

$$\left| \sum_{k \leq 2n} c_k - \left(\sum_{k \leq n} a_k\right)\left(\sum_{k \leq n} b_k\right) \right| \leq R_0 T_{n+1} + T_0 R_{n+1},$$

since the terms that remain on the LHS after cancellation are of the form $a_k b_l$ where either $k \geq n+1$ or $l \geq n+1$. Upon taking the limit as $n \longrightarrow \infty$, we obtain (1.45). [18] ♠

An important property of an absolutely convergent series is:

[18] see Exercise 4 at the end of this section.

Theorem 1.4.6 *Let $\sum_n z_n$ be an absolutely convergent series. Then every rearrangement $\sum_n z_{\sigma_n}$ of the series is also absolutely convergent.*

Recall that a rearrangement $\sum_n z_{\sigma_n}$ of $\sum_n z_n$ is obtained by taking a bijection $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$.

We shall now consider the study of a family of sequences. Let $f_n : A \to \mathbb{C}$ be a sequence of functions taking complex values defined on some set A. Then to each $x \in A$, we get a sequence $\{f_n(x)\}$. If each of these sequences is convergent, then we get a function $f(x) = \lim_{n\to\infty} f_n(x)$. In the definition of the convergence, it is obvious that the largeness of n_0 depends on the smallness of ϵ. Naturally, it would also vary from sequence to sequence. If we want the function f to behave well, it is clearly necessary that there must be some control on the variation of n_0 from sequence to sequence. One way to control this leads us to the notion of uniform convergence.

Definition 1.4.6 Let $\{f_n\}$ be a sequence of complex valued functions on a set A. We say that it is *uniformly convergent* on A to a function f if for every $\epsilon > 0$ there exists n_0, such that for all $n \geq n_0$, we have, $|f_n(x) - f(x)| < \epsilon$, for all $x \in A$.

Remark 1.4.5 Observe that if $\{f_n\}$ is uniformly convergent on A, then for each $x \in A$, we have, $f_n(x) \longrightarrow f(x)$. This is called *point-wise convergence* of the sequence of functions. As seen in the example below, point-wise convergence does not imply uniform convergence. However, it is fairly easy to see that this is so if A is a finite set. Thus the interesting case of uniform convergence occurs only when A is an infinite set. The terminology is also adopted in an obvious way for series of functions via the associated sequences of partial sums. As in the case of ordinary convergence, we have Cauchy's criterion here also.

Theorem 1.4.7 *A sequence of complex valued functions $\{f_n\}$ is uniformly convergent iff it is* uniformly Cauchy *i.e., given $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0, p \geq 0$ and for all $x \in A$, we have,*

$$|f_{n+p}(x) - f_n(x)| < \epsilon.$$

Example 1.4.1 A simple example of a sequence which is point-wise convergent but not uniformly convergent is $f_n : \mathbb{R}^+ \longrightarrow \mathbb{R}$ given by $f_n(x) = 1/nx$. It is uniformly convergent in $[\alpha, \infty)$ for all $\alpha > 0$ but not so in $(0, \alpha)$.

Example 1.4.2 The mother of all convergent series is the geometric series

$$1 + z + z^2 + \cdots$$

The sequence of partial sums is given by

$$1 + z + \cdots + z^{n-1} = \frac{1 - z^n}{1 - z}.$$

For $|z| < 1$ upon taking the limit we obtain

$$\boxed{1 + z + z_2 + \cdots + z^n + \cdots = \frac{1}{1 - z}.} \tag{1.46}$$

In fact, if we take $0 < r < 1$, then in the disc $B_r(0)$, the series is uniformly convergent. For, given $\epsilon > 0$, choose n_0 such that $r^{n_0} < \epsilon(1 - r)$. Then for all $|z| < r$ and $n \geq n_0$, we have,

$$\left|\frac{1 - z^n}{1 - z} - \frac{1}{1 - z}\right| = \left|\frac{z^n}{1 - z}\right| \leq \frac{|z^{n_0}|}{1 - |z|} < \epsilon$$

There is a pattern in what we saw in the above example. This is extremely useful in determining uniform convergence:

Theorem 1.4.8 Weierstrass[19]M-test: *Let $\sum_n a_n$ be a convergent series of positive terms. Suppose there exists $M > 0$ and an integer N such that $|f_n(x)| < Ma_n$ for all $n \geq N$ and for all $x \in A$. Then $\sum_n f_n$ is uniformly and absolutely convergent in A.*

Proof: Given $\epsilon > 0$ choose $n_0 > N$ such that $a_n + a_{n+1} + \cdots + a_{n+p} < \epsilon/M$, for all $n \geq n_0$. This is possible by Cauchy's criterion, since $\sum_n a_n$ is convergent. Then it follows that

$$|f_n(x)| + \cdots + |f_{n+p}(x)| \leq M(a_n + \cdots + a_{n+p}) < \epsilon,$$

for all $n \geq n_0$ and for all $x \in A$. Again, by Cauchy's criterion, this means that $\sum f_n$ is uniformly and absolutely convergent. ♠

Remark 1.4.6 The series $\sum_n a_n$ in the above theorem is called a 'majorant' for the series $\sum_n f_n$. Here is an illustration of the importance of uniform convergence. (See definition 1.5.1 for continuous functions.)

[19]Karl Weierstrass (1815-1897) a German mathematician is well known for his perfect rigor. He clarified any remaining ambiguities in the notion of a function, of derivatives, of minimum etc., still prevalent in his time.

Theorem 1.4.9 *Let $\{f_n\}$ be a sequence of continuous functions defined and uniformly convergent on a subset A of $\mathbb{R}$ or $\mathbb{C}$. Then the limit function $f(x) = \lim_{n\longrightarrow\infty} f_n(x)$ is continuous on A.*

Proof: Let $x \in A$ be any point. In order to prove the continuity of f at x, given $\epsilon > 0$ we should find $\delta > 0$ such that for all $y \in A$ with $|y-x| < \delta$, we have, $|f(y)-f(x)| < \epsilon$. So, by the uniform convergence, first we get n_0 such that $|f_{n_0}(y) - f(y)| < \epsilon/3$ for all $y \in A$. Since f_{n_0} is continuous at x, we also get $\delta > 0$ such that for all $y \in A$ with $|y-x| < \delta$, we have $|f_{n_0}(y) - f_{n_0}(x)| < \epsilon/3$. Now, using triangle inequality, we get,

$$|f(y) - f(x)| \leq |f(y) - f_{n_0}(y)| + |f_{n_0}(y) - f_{n_0}(x)| + |f_{n_0}(x) - f(x)| < \epsilon,$$

whenever $y \in A$ is such that $|y-x| < \delta$. ♠

Exercise 1.4

1. Let $z_n = x_n + \imath y_n, n \geq 1$. Show that $z_n \to z = x + \imath y$ iff $x_n \to x$ and $y_n \to y$.

2. Let $\sum_n a_n$ be a convergent series of non negative real numbers. Show that $\sum_n a_n^2$ is convergent.

3. Let $\sum_n z_n$ be a convergent series of complex numbers such that $\Re(z_n) \geq 0$ for all n. If $\sum_n z_n^2$ is also convergent, show that $\sum_n |z_n|^2$ is convergent.

4. If a sequence converges to a limit, then every subsequence of it converges to the same limit. This was implicitly used in the proof of theorem 1.4.5. Pin-point where exactly we did this.

5. Give an example of a sequence of complex numbers z_n which converges to a complex number $z \neq 0$, yet $\operatorname{Arg} z_n$ does not converge to $\operatorname{Arg} z$. However, if $z_n \longrightarrow z(\neq 0)$ then show that there exist $\theta_n \in \arg\, z_n$ such that $\theta_n \longrightarrow \theta \in arg\, z$.

6. For $0 \leq \theta < 2\pi$ and for any $\alpha \in \mathbb{R}$, define the *closed sector $S(\alpha,\theta)$ with span θ* by

$$S(\alpha,\theta) = \{rE(\beta) \ : \ r \geq 0 \,\&\, \alpha \leq \beta \leq \alpha + \theta\}.$$

   Let $\sum_n z_n$ be a convergent series. If $z_n \in S(\alpha,\theta), n \geq 1$, where $\theta < \pi$, then show that $\sum_n |z_n|$ is convergent. (This is an improvement on Exercise 3 above!)

7. Let $\sum_n z_n$ be a series of complex numbers so that each of its four subseries consisting of terms lying in the same closed quadrant is convergent. Show that $\sum_n |z_n|$ is convergent.

8. **Telescoping:** Given a sequence $\{x_n\}$ define the difference sequence $a_n := x_n - x_{n+1}$. Then show that the series $\sum_n a_n$ is convergent iff the sequence $\{x_n\}$ is convergent and in that case, $\displaystyle\sum_n a_n = x_0 - \lim_{n \longrightarrow \infty} x_n$. Employ this to compute $\displaystyle\sum_n \frac{1}{n(n+1)}$.

9. Let $\{z_n\}$ be a bounded sequence and $\sum_n w_n$ is an absolutely convergent series. Show that $\sum_n z_n w_n$ is absolutely convergent.

10. **Abel's Test:** For any sequence of complex numbers $\{a_n\}$, define $S_0 = 0$ and $S_n = \sum_{k=1}^n a_k,\ n \geq 1$. Let $\{b_n\}$ be any sequence of complex numbers.
(i) Prove **Abels' Identity**:
$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n-1} S_k(b_k - b_{k+1}) - S_{m-1}b_m + S_n b_n,\ 1 \leq m \leq n.$$
(ii) Show that $\sum_n a_n b_n$ is convergent if the series $\sum_k S_k(b_k - b_{k+1})$ is convergent and $\displaystyle\lim_{n \longrightarrow \infty} S_n b_n$ exits.
(iii) **Abel's Test:** Let $\sum_n a_n$ be a convergent series and $\{b_n\}$ be a bounded monotonic sequence of real numbers. Then show that $\sum_n a_n b_n$ is convergent.

11. **Dirichlet's Test:** Let $\sum_n a_n$ be such that the partial sums are bounded and let $\{b_n\}$ be a monotonic sequence tending to zero. Then show that $\sum_n a_n b_n$ is convergent.

12. Derive the following **Leibniz's test** from Dirichlet's Test: If $\{a_n\}$ is a monotonic sequence converging to 0 then the alternating series $\sum_n (-1)^n a_n$ is convergent.

13. Generalize the Leibniz's test as follows: If $\{a_n\}$ is a monotonic sequence converging to 0 and $|\zeta| = 1$, $\zeta \neq 1$, then $\sum_n \zeta^n a_n$ is convergent.

14. Show that if $\sum_n a_n$ is convergent then the following sequences are all convergent.
(a) $\displaystyle\sum_n \frac{a_n}{n^p}, p > 0$; (b) $\displaystyle\sum_n \frac{a_n}{log_p n}$; (c) $\sum_n \sqrt[n]{n} a_n$; (d) $\displaystyle\sum_n \left(1 + \frac{1}{n}\right)^n a_n$.

15. Show that for any $p > 0$, and for every real number x, $\displaystyle\sum_n \frac{\sin nx}{n^p}$ is convergent.

16. Consider the series $\sum_n \frac{z^n}{1-z^n}$. Determine the domain on which the sum defines a continuous function.

1.5 Topological Aspects of Complex Numbers

The availability of the concept of *'absolute value'* which obeys properties such as triangle inequality enables us to introduce the concept of a *'distance function'* on $\mathbb{C}$, making it into a metric space. The study of the metric aspects of $\mathbb{C}$ can be caried out, without any extra effort in a slightly more general set-up of Euclidean spaces of all dimension simultaneously and that is what we are going to do now.

Thanks to Descartes[20], we can refer to each and every corner of our space, in a very precise manner on our finger tips! For $n \geq 2$, the n-dimensional Cartesian coordinate space is the totality of all ordered n-tuple $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ of real numbers $x_j \in \mathbb{R}$. It is denoted by $\mathbb{R} \times \cdots \times \mathbb{R}$ (n factors) or, in short by $\mathbb{R}^n$. Elements of this space are often referred to as 'n-vectors'. For $n = 1$ we have $\mathbb{R}^1 = \mathbb{R}$. Observe that every real number is a 1-vector. For n=2 and 3, we can avail of the familiar notation $(x, y) \in \mathbb{R}^2$ and $(x, y, z) \in \mathbb{R}^3$ etc., respectively. For a point $\mathbf{x} = (x_1, \ldots, x_n)$, we refer to x_j as the j-th *coordinate* of $\mathbf{x}$. For each $j = 1, 2, \ldots, n$, let $\pi_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ denote the function defined by $\pi_j(x_1, \ldots, x_n) = x_j$. It is called the j^{th} *coordinate function* or the j^{th} *coordinate projection.* These are the first important examples of functions from $\mathbb{R}^n$ to $\mathbb{R}$.

Given any function $f : A \longrightarrow \mathbb{R}^n$, where A is any set, we can compose it with the j^{th} coordinate projection to get n functions $f_j := \pi_j \circ f$. An important observation is that f is completely determined by these functions f_j called the j^{th} *component function of* f, viz., $f(a) = (f_1(a), \ldots, f_n(a))$, $\forall\, a \in A$. Thus one can expect that the study of functions $f : A \longrightarrow \mathbb{R}^n$ can be effectively reduced to the study of functions $f_j : A \longrightarrow \mathbb{R}$, n of them in number, and perhaps to the study of their inter-relation. Soon we shall see that this is in fact so, to a large extent.

Let us now discuss two major, closely related structural aspects on $\mathbb{R}^n$. The first one is the algebraic structure and the second, geometric. Some of the algebraic structure of the real numbers induce similar structures on $\mathbb{R}^n$. The most important amongst them is the *'sum' or 'addition'.* Given two n-vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, we define their sum:

$$\mathbf{x} + \mathbf{y} := (x_1 + y_1, \ldots, x_n + y_n).$$

It is straightforward to verify that this operation obeys all the usual laws that the addition of real numbers obeys. In particular, observe that the zero element for this sum is the zero vector $\mathbf{0} = (0, \ldots, 0)$ and the negative of $(x_1, \ldots, x_n)$ is $(-x_1, \ldots, -x_n)$.

[20]René Descartes (1596-1650) was a French philosopher-mathematician.

Next, for each real number α and a vector $\mathbf{x}$, we define

$$\alpha\mathbf{x} = (\alpha x_1, \ldots, \alpha x_n).$$

It has the following usual properties:
(V1) *Associativity*: $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$;
(V2) *Distributivity over the sum*: $\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x}+\alpha\mathbf{y}$, and
(V3) *Identity*: $1\mathbf{x} = \mathbf{x}$.

This structure is what makes $\mathbb{R}^n$ into a n-dimensional vector space over $\mathbb{R}$. In particular, we also observe that, if $n = 1$, then these two operations coincide with the usual addition and multiplication.

For any two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ as above, we define the distance between $\mathbf{x}$ and $\mathbf{y}$ by

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2}.$$

Indeed, the distance function, in this case, arises out of the dot product

$$(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}\cdot\mathbf{y} := \sum_j x_j y_j.$$

One defines the *norm* function by $\|\mathbf{x}\| := \sqrt{\sum_{j=1}^n x_j^2} = \sqrt{\mathbf{x}\cdot\mathbf{x}}$. Clearly, $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$.

You may be familiar with various *metric* properties of this distance function such as
(M1) *symmetry*: $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$;
(M2) *triangle inequality*: $d(\mathbf{x}, \mathbf{y}) \le d(\mathbf{x}, \mathbf{w}) + d(\mathbf{w}, \mathbf{y})$;
(M3) *positivity*: $d(\mathbf{x}, \mathbf{y}) \ge 0$ and $= 0$ if and only if $\mathbf{x} = \mathbf{y}$. Moreover, we also have,
(M4) *homogeneity*: $d(\alpha\mathbf{x}, \alpha\mathbf{y}) = |\alpha| d(\mathbf{x}, \mathbf{y})$.

All these properties are standard and can be verified easily. For instance, in the proof of triangle inequality, we can use the famous
(M5) *Cauchy-Schwarz*[21] *inequality*: $|\mathbf{x}\cdot\mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\|$.

[The inequality above can be deduced from Cauchy's inequality (1.12) by putting $z_j = x_j$ and $w_j = y_j$. Alternatively, use the fact $(x_jT + y_j)^2 \ge 0$ for all j and for $T \in \mathbb{R}$, sum it up over j and then appeal to the discriminant criterion of a quadratic aT^2+bT+c to take only non negative values.]

[21] H.A. Schwarz (1843-1921) was born in Hernsdorf, Poland, (which is now in Germany). He worked on the conformal mapping of polyhedral surfaces onto the spherical surface, minimal surfaces, Dirichlet's problem etc.. Several important results and concepts in Analysis bear his name.

The n-dimensional real Cartesian coordinate space together with the above vector space and metric structure is called the *n-dimensional Euclidean space.*

The distance function allows us to talk about 'nearness' of points. In particular, the concept of limits of sequences, Cauchy-sequences, continuity of functions etc., can be introduced just as in $\mathbb{R}$. It is then easily verified that a sequence of complex numbers converges iff the corresponding sequences of real and imaginary parts both converge.

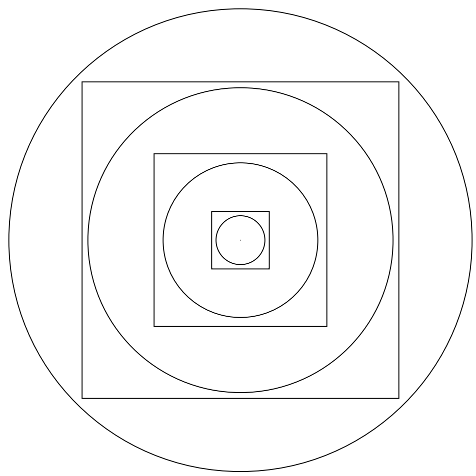


Fig. 7

The crux of the matter is the following principle:

> The sequence $\{\mathbf{x}_k\}$ has a particular property iff the corresponding n coordinate sequences $\{\pi_j(\mathbf{x}_k)\}$ all have the same property.

The underlying fact that yields this principle can be stated as follows:

> Inside each ball with a given center there is a rectangular box with the same center and vice versa.

This is depicted in Fig. 7.

Recall that a function $f : (a, b) \to \mathbb{R}$ is defined to be continuous at $t_0 \in (a, b)$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $t \in (a, b) \cap (t_0 - \delta, t_0 + \delta)$, we have,

$$|f(t) - f(t_0)| < \epsilon.$$

The function f is said to be continuous on (a, b) if it is so at every point of (a, b).

Following the same line, let us generalize the definition of continuity a little bit:

Definition 1.5.1 Given a subset $X \subset \mathbb{R}^n$ and a function $f : X \to \mathbb{R}^m$, we say f is continuous at $z \in X$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $w \in X$, with $\|w - z\| < \delta$ we have, $\|f(w) - f(z)\| \leq \epsilon$.

You can now verify in a straight forward manner that the projection maps $\pi_j : \mathbb{R}^m \to \mathbb{R}$ are all continuous and a map $f : X \to \mathbb{R}^m$ is continuous iff each $\pi_j \circ f$ is continuous. Similar property holds for real differentiability as well. (See section 3.2.)

For $n < m$, we can think of $\mathbb{R}^n$ as a subset of $\mathbb{R}^m$ in various ways. One such standard way is to identify it with the set of points $\mathbf{y}$ of $\mathbb{R}^m$ whose last $m - n$ coordinates are identically zero. Indeed, the *'inclusion map'* $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, 0, \ldots 0)$ defines such an identification. Observe that, more generally, we could also have chosen any other $m-n$ coordinates to vanish identically, to get other inclusion maps. In particular, for $n = 1$, we get the m *coordinate axes* of $\mathbb{R}^m$ given by the inclusions:

$$x \mapsto (0, \ldots, 0, x, 0, \ldots, 0)$$

the entry x being taken in the j^{th} place, for $j = 1, 2, \ldots, m$.

However, compared with projection maps, the inclusion maps are much weaker in capturing the behavior of a given function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. For instance, there are discontinuous functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that for every fixed x_0, y_0, the functions $x \mapsto f(x, y_0)$ and $y \mapsto f(x_0, y)$ are continuous. Similar statement holds for differentiability also. (We shall not go into more details of this here. See the exercises.) This weakness of the inclusion maps is precisely what makes the calculus of several variables, not a mere extension of calculus of 1-variable but something deeper and interesting. Nevertheless, it is possible to extract a large amount of information on the function by studying it after restricting it to various line segments. In any case, first of all, we need to study carefully the 'topological structure'[22] of $\mathbb{R}^n$ arising out of the distance function.

We shall consider subsets of some Euclidean space and refer to them as 'spaces'. Also in this section, we shall consider functions defined on subsets of some Euclidean space.

In the case of real numbers, the basic domains of definitions of differentiable functions were open intervals. In $\mathbb{R}^n$ the prototype of open intervals are the open balls. For any

[22]A topological structure on an arbitrary set X is a collection of subsets of X to be called 'open subsets' of X which includes the empty set and the whole set and satisfies the two conditions (i) and (ii) above. Whenever, a set is given such a structure, it is called a topological space. A metric space is always given the structure of a topological space where open sets are those which are union of open balls.

$\mathbf{x} \in \mathbb{R}^n$ and $r > 0$, let

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \ : \ d(\mathbf{x}, \mathbf{y}) < r\}; \ \ \overline{B}_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \ : \ d(\mathbf{x}, \mathbf{y}) \leq r\}. \tag{1.47}$$

These subsets of $\mathbb{R}^n$ are called the *open* (respectively, *closed*) *balls of radius* r *around* $\mathbf{x}$. For the study of differentiation, the functions should be defined on an open ball around a point. This automatically leads us to the concept of **open sets**:

Definition 1.5.2 A subset U of $\mathbb{R}^n$ is called open if it is the union of a family of open balls.

Remark 1.5.1 Observe that a subset $A \subset \mathbb{R}^n$ is *open* if for each $\mathbf{x} \in A, \ \exists \ r > 0$ such that $B_r(\mathbf{x}) \subset A$. Moreover, every open set is the union of all the open balls that are contained in it. It is not hard to see that:
(i) if A and B are open then $A \cap B$ is also open;
(ii) union of any family of open sets in $\mathbb{R}^n$ is open.
(iii) The whole space $\mathbb{R}^n$ is open.
(iv) The empty set $\emptyset$ is open because the condition is satisfied vacuously!

These properties of open sets are the most fundamental properties and constitute the definition of a *topological structure* on $\mathbb{R}^n$. Then each subset X of $\mathbb{R}^n$ also inherits a topological structure as follows: A subset U of X is said to be *open in* X if there is an open subset V of $\mathbb{R}^n$ such that $U = V \cap X$. One can easily verify that the family of open sets in X obeys the two laws specified above.

Definition 1.5.3 We say a subset F of X is *closed* in X if the set theoretic complement $X \setminus F$ is open.

Remark 1.5.2 Note that a subset of X may happen to be closed as well as open in X or neither. We shall not go into more details here. However, the following characterization of a closed subset of $\mathbb{R}^n$ is quite useful and indeed will be used in this text very often.

Theorem 1.5.1 *A subset F of $\mathbb{R}^n$ is closed in $\mathbb{R}^n$ iff for every convergent sequence of points in F the limit of the sequence also is in F.*

As a simple illustration of topological methods, let us have some alternative definitions of continuity.

Theorem 1.5.2 *Let $f : X \longrightarrow Y$ be any function. Then the following statements are all equivalent to each other.*
(i) *f is continuous.*
(ii) *For each open subset V of Y, $f^{-1}(V)$ is open in X.*
(iii) *For each closed subset F of Y, $f^{-1}(F)$ is closed in X.*

Proof: (i) $\Longrightarrow$ (ii): Let f be continuous. Let V be open in Y. To show that $f^{-1}(V)$ is open in X, let $z_0 \in f^{-1}(V)$. This means $f(z_0) \in V$. Since V is open, there exists $\epsilon > 0$ such that $|w - f(z_0)| < \epsilon$ and $w \in Y$ implies $w \in V$. Then by continuity of f, there exists $\delta > 0$ such that $|z - z_0| < \delta$ and $z \in X$ implies that $|f(z) - f(z_0)| < \epsilon$ and hence $f(z) \in V$. This, in turn means $z \in f^{-1}(V)$. Therefore $f^{-1}(V)$ is open.
(ii) $\Longrightarrow$ (i): Fix $z_0 \in X$ and let us show that f is continuous at z_0. Given $\epsilon > 0$, the set $V = B_\epsilon(f(z_0)) \cap Y$ is an open subset of Y and $f^{-1}(V)$ is open in X. Since $z_0 \in f^{-1}(V)$, there exists $\delta > 0$ such that $B_\delta(z_0) \subseteq f^{-1}(V)$. This means that $f(B_\delta(z_0) \cap X) \subset V \subset B_\epsilon(f(z_0))$. This is the same as saying that for all $z \in X$ with $|z - z_0| < \delta$, we have, $|f(z) - f(z_0)| < \epsilon$.
(ii) $\Longleftrightarrow$ (iii): This follows by a purely set-theoretic property of taking inverses and complements: for any subset $B \subset Y$ we have $X \setminus f^{-1}(B) = f^{-1}(Y \setminus B)$.

♠

One of the important topological concepts is the notion of *compactness.* In the case of subsets of $\mathbb{R}^n$ we could take the following tentative definition of compactness. The general definition is quite different and we do not need it for quite some time.

Definition 1.5.4 A subset of $\mathbb{R}^n$ is called *compact* iff it is closed and bounded.

The following are a few important properties of compact subsets (of $\mathbb{R}^n$). We shall not give any proofs of these here. However, because of their importance, for the sake of completeness of the exposition, we indicate proofs in the exercises below.

Theorem 1.5.3 Bolzano[23]-Weierstrass Property *Let A be a bounded subset of $\mathbb{R}^n$. Then every infinite sequence in A has a subsequence that is convergent to a point in $\mathbb{R}^n$. Further if A is a closed subset then this limit point belongs to A.*

Theorem 1.5.4 Uniform Continuity *Every continuous real valued function on a compact subset of $\mathbb{R}^n$ is uniformly continuous.*

[23]Bernard Bolzano (1781-1848) an Austrian mathematician and professor of religious studies is known for his studies in point sets and foundational mathematics.

Theorem 1.5.5 Attainment of Sup-Inf *Every continuous real valued function on a compact subset K of $\mathbb{R}^n$ attains its supremum and infimum on K.*

Theorem 1.5.6 Intermediate Value Property (IVP) *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then $f([a, b])$ is an interval.*

We shall wind up this section with a discussion of a profound yet extremely simple property of open sets.

Consider an open ball in any Euclidean space. Observe that the norm function does not attain its supremum on it. Indeed, none of the coordinate functions attain their supremum or infimum on it. It follows that the same is true for any open set in the Euclidean space. This phenomenon is the essence of the celebrated *maximum modulus principle* for analytic functions and the so called *maximum and minimum principle* for harmonic functions, that we are going to study later in this course. So, it is good to become familiar with this notion at an early stage. However, you may entirely skip this part if you feel that you are not up to it.

Let us first introduce a definition and then state the observations that we have made above as a theorem.

Definition 1.5.5 Let $f : X \longrightarrow Y$ be a function, where X, Y are subspaces of some Euclidean spaces. We say f is an *open mapping* if each open subset of X is mapped onto an open subset of Y.

Let us introduce the notation $\|f\|$ to denote the function $\mathbf{x} \mapsto \|f(\mathbf{x})\|$.

Theorem 1.5.7 *Let V be an open subset of $\mathbb{R}^n$ and $f : V \longrightarrow \mathbb{R}^m$ be an open mapping. Let U be any non empty open subset of V. Then the following are true.*
(i) *For $m = 1$, f neither attains its supremum nor attains its infimum on U.*
(ii) *None of the coordinate functions $f_j := \pi_j \circ f$ attains its supremum or infimum on U.*
(iii) *$\|f\|$ does not attain its supremum on U.*
(iv) *If f never vanishes on U then $\|f\|$ does not attain its infimum on U.*

Proof: (i) Since U is open and f is an open mapping, $f(U)$ is open in $\mathbb{R}$. This means that given any point $\mathbf{x} \in U$, there exists $\epsilon > 0$ such that the open interval $(f(x) - \epsilon, f(x) + \epsilon) \subset f(U)$. Therefore $f(x)$ is neither the maximum nor the minimum on U. Since this is true of all points $\mathbf{x} \in U$, statement (i) follows.
(ii) We just observe that the coordinate functions π_j are open mappings and composite of two

open mappings is again an open mapping. Now use (i).

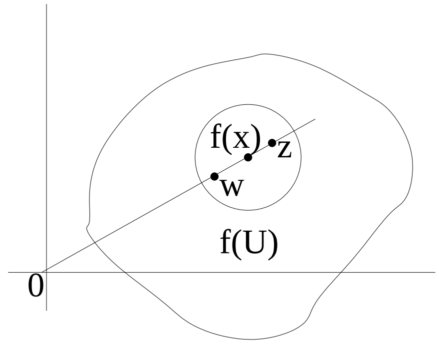


Fig. 8

(iii) Since $f(U)$ is open, for a given $\mathbf{x} \in U$, we can find $\epsilon > 0$ such that $B_\epsilon(f(\mathbf{x})) \subset f(U)$. But then this open ball contains points z which are farther than $f(\mathbf{x})$ from the origin. (See the Fig. 8). This is clear if $f(\mathbf{x}) = 0$; otherwise, we can take $\bar{z} = \dfrac{2\|f(\mathbf{x})\| + \epsilon}{2\|f(x)\|} f(\mathbf{x})$, so that $\|\bar{z}\| > \|f(\mathbf{x})\|$. Since $\|\bar{z} - f(\mathbf{x})\| = \epsilon/2$, we have $\bar{z} \in f(U)$ and This proves (iii).
(iv) When $f(\mathbf{x})$ never vanishes, put $\bar{z} = \dfrac{2\|f(\mathbf{x})\| - \epsilon}{2\|f(x)\|} f(\mathbf{x})$ so that $\|\bar{z}\| < \|f(\mathbf{x})\|$. Again, since $\|\bar{z} - f(\mathbf{x})\| = \epsilon/2$, $z \in f(U)$. This proves (iv). ♠

Remark 1.5.3 Observe that we cannot say that $\|f\|$ does not attain its infimum, in general. The absolute minimum for the modulus function being 0, the moment $0 \in f(U)$, we are done.

Another important topological notion viz., path connectivity will be developed in the next section.

Exercise 1.5

1. Show that a complex valued function f is continuous iff the two functions obtained by taking the real and imaginary parts of f are continuous.

2. Show that the following subsets of the complex plane are open:
   (a) $\mathbb{C} \setminus A$, where A is a finite subset of $\mathbb{C}$;

   (b) $\{z \;:\; |z| > r\}$, where $r \in \mathbb{R}$;

   (c) $\{x + \imath y \;:\; x \neq 0\}$;

   (d) $\{x + \imath y \;:\; a < x < b,\ c < y < d\}, a, b, c, d \in \mathbb{R}$;

3. For a continuous function $f : \mathbb{C} \longrightarrow \mathbb{R}$, directly verify that $f^{-1}(-1, 1)$ is an open subset of $\mathbb{C}$, without quoting theorem 1.5.2.

4. Let $f, g : \mathbb{C} \longrightarrow \mathbb{C}$ be continuous functions. Show that $\{z : f(z) = g(z)\}$ is a closed subset of $\mathbb{C}$.

5. Show that none of the following subsets of $\mathbb{R}^2$ are open in it:

   (i) A line $ax + by = c = 0$; (ii) a circle $(x - a)^2 + (y - b)^2 = k^2$;
   (iii) The boundary of a convex polygon; (iv) any countable subset;
   (v) The closed disc $\{z : |z - a| \leq r\}$.

6. X be a totally ordered set i.e., one with an order satisfyning FVII. Show that every sequence in X has a subsequence which is monotone. (Indeed prove this for a sequence of rational numbers without using the existence of real numbers. Then imitate the same proof in the general case so that you get proof which does not use the 'completeness' property (1.14) or any of its consequences.)

7. Show that every bounded sequence of real numbers has a subsequence which is convergent. [Hint: Use (1.14)]

8. Use theorem 7 above to show that every Cauchy sequence in $\mathbb{R}$ is convergent.

9. * Deduce the general theorem 1.5.3 from the case when $n = 1$ as in the above exercise 7.

10. * Use theorem 1.5.3 along with theorem 1.4.2 to prove theorem 1.5.4.

11. * Show that image of a compact set under a continuous function is compact.

12. * Use Ex. 11 to give a proof of theorem 1.5.5.

13. * Show that if U is a non empty subset of $[a, b]$ which is both open and closed, then $U = [a, b]$.

14. * Prove theorem 1.5.6.

15. Combine IVP with the Archimedian property, to prove the following: for any positive integer the map $p : [0, \infty) \to [0, \infty)$ given by $p(r) = r^k$ is surjective.

1.6 Path Connectivity

Definition 1.6.1 Let A be a subset of the complex plane. By a *path, a curve* or an *arc* in A, we mean a continuous map $\gamma : [a, b] \longrightarrow A$.

The points $\gamma(a)$ and $\gamma(b)$ are called the *end points* of γ. In fact $\gamma(a)$ is called the *initial point* and $\gamma(b)$ is called the *terminal point.* We also say that γ is a path joining $z_1 := \gamma(a)$ and $z_2 := \gamma(b)$. If such a path exists in A, we say that z_1 and z_2 can be *joined in A*.

Given two paths $\gamma_i : [a_i, b_i] \longrightarrow A,\ i = 1, 2$, suppose that $\gamma_1(b_1) = \gamma_2(a_2)$. Then we define the composite path $\gamma := \gamma_1 \cdot \gamma_2 : [a, b] \longrightarrow A$ by taking $a = a_1,\ b = b_1 + b_2 - a_2$ and

$$\gamma(t) = \begin{cases} \gamma_1(t), & \text{for } a_1 \leq t \leq b_1; \\ \gamma_2(t + a_2 - b_1), & \text{for } b_1 \leq t \leq b_1 + b_2 - a_2. \end{cases} \tag{1.48}$$

By the *inverse path* γ^{-1} of a given path $\gamma : [a, b] \longrightarrow A$ we mean the path defined by $\gamma^{-1}(t) = \gamma(a + b - t),\ a \leq t \leq b$. It is indeed the path γ traversed in the 'opposite direction'.

Remark 1.6.1

1. Often, by a *curve* one means the image of a path γ, and γ itself is referred to as a *parameterization* of the curve. For instance, the circle $\{z : |z| = 1\}$ is thought of as a curve and then $\gamma(\theta) = (\cos\theta, \sin\theta)$ [or equivalently $\gamma(\theta) := \cos\theta + \imath \sin\theta$], $0 \leq \theta \leq 2\pi$, is thought of as a parameterization of the circle. We shall denote the image of γ by $Im(\gamma)$.

2. Observe that since $\gamma_2(b_1 + a_2 - b_1) = \gamma_2(a_2) = \gamma_1(b_1)$, it follows that γ as given by (1.48) is well defined and continuous. Thus we see that the composite path is obtained by first traversing along γ_1 and then along γ_2. Clearly $Im(\gamma) = Im(\gamma_1) \cup Im(\gamma_2)$.

3. Also note that if $\gamma : [a, b] \longrightarrow A$ is a path and $a = a_1 < a_2 < a_3 = b$, is a subdivision then indeed, $\gamma = \gamma_1.\gamma_2$, where γ_i are the restrictions of γ to the two sub-intervals $[a_1, a_2]$ and $[a_2, a_3]$. This remark will be of crucial practical importance to us soon.

It is important that we should be able to move from one point to another continuously, remaining within a specified set. This leads us to the notion of path connectedness.

Definition 1.6.2 We say A is *path connected* if any two points $z_1, z_2 \in A$ can be joined in A, i.e., we can find a path within A with end points as z_1 and z_2.

Remark 1.6.2 If we have a point z_0 to which every other point of A can be joined in A then given any two points z_1 and z_2, we can take a path from z_1 to z_0 and then follow it by a path from z_0 to z_2. Hence, A is path connected. Obvious examples of (path) connected spaces are singleton sets. Of course any convex subset of $\mathbb{C}$ is path connected. (Recall that a subset A of $\mathbb{C}$ is said to be *convex* if $z_1, z_2 \in A$ implies the entire line segment $[z_1, z_2] = \{(1-t)z_1 + tz_2, 0 \le t \le 1\}$ is contained in A.)

Remark 1.6.3 There is a 'method' in which path connectedness is exploited very often. To illustrate this, we shall first introduce the following concept and then prove a theorem.

Definition 1.6.3 Let X be a topological space and $f : X \longrightarrow Y$ be a function, where Y is any set. We say f is *locally constant* if for each $x \in X$, there exists an open set U_x in X such that $x \in U_x$ and for all $y \in U_x$, we have $f(y) = f(x)$.

It follows that the inverse image of a singleton set under such a function f is an open set. Indeed, even the converse is true and we can say that f is a locally constant function iff the inverse image of every set is open. Observe that we have made no assumptions on the co-domain Y of the function f in this definition. Thus, if Y happens to be any topological space, then it turns out that f will be automatically a continuous function.

Theorem 1.6.1 *A locally constant function on a path-connected space is a constant.*

Proof: Let f be a locally constant function on X. Fix a point $a \in X$ and define $\alpha : X \to [0,1]$ by the property that

$$\alpha(z) = \begin{cases} 0, & \text{if } f(z) = f(a); \\ 1, & \text{otherwise .} \end{cases}$$

Then verify that α is a continuous function. Now let $b \in X$ be any other point and $\gamma : [0,1] \to X$ be a path such that $\gamma(0) = a, \gamma(1) = b$. Then $\alpha \circ \gamma : [0,1] \to [0,1]$ is a continuous map which takes only two possible values 0 or 1. Therefore it should take only one value, viz. 0. This implies $\alpha(b) = 0$ and hence $f(b) = f(a)$. ♠

As a useful corollary we have:

Theorem 1.6.2 *Let Ω be an open path connected subset of $\mathbb{C}$. If A is a non empty subset of Ω which is both open and closed then $A = \Omega$.*

Proof: If not, we can consider the function $\chi : \Omega \to [0,1]$ such that $\chi(z) = 1$ if $z \in A$ and $= 0$ if $z \notin A$. Then χ will be a locally constant function which is not a constant. ♠

Remark 1.6.4

1. The 'method' that we alluded above is hidden in the above theorem. On a space X which is path connected, suppose we want to prove a certain topological property $\mathbf{P}$ at all points of X. We consider the subset A of all such points of X at which $\mathbf{P}$ is true. We then show that A is non empty, open and closed in X. The theorem then says that $A = X$. We shall have at least some opportunity to see this method in action.

2. There is a notion of 'connectivity' which is somewhat weaker and less intuitive than path-connectivity, which we shall introduce in the next section. However, the notions coincide for open subsets of $\mathbb{C}$. Thus it turns out that you need not worry about it except for the sake of one small result in chapter 7. Until then **we shall use the word 'connected' to mean 'path connected'** just to save that much of breath and space.

3. The concept of path connectivity is very important everywhere in science. It means that communications of any kind whatsoever between any two points is possible.

4. It follows easily that being 'joined by an arc in A' is an equivalence relation on the set of points of A. Therefore, A will get partitioned into equivalence classes called the *path-components* or in our new convention, *components of* A. A component of A can also be described as a maximal connected subset of A. Often the study of topological properties of a space is restricted to one such component at a time. This method allows us to make a technical assumption that the space A is connected.

5. The plane $\mathbb{C}$ itself is connected. Indeed, every *convex* subset of a Euclidean space is connected. In particular closed and open discs are connected.

6. More generally, a subset $A \subset \mathbb{R}^n$ is called *star-shaped* if there exists $a \in A$ such that for all $b \in A$ we have the line segment $[a, b] \subseteq A$. It follows that every star-shaped subset of $\mathbb{R}^n$ is connected.

7. Try to prove that even if you delete a finite number of points from $\mathbb{C}$ it remains connected. Then try to improve upon this result by removing an infinite but countable subset.

Definition 1.6.4 By a *region* or a *domain* in $\mathbb{C}$ we mean an open subset of $\mathbb{C}$ which is connected.

The reason for over-using the word *domain* in this special sense is that it is already widely used in the literature. In classical analysis, the domain (in a set theoretic sense) of definition of a function is always restricted to this type of subsets, viz., those which are *open* and *connected.* Openness is often necessary to carry out differentiation. The connectivity assumption can be justified by results of the following type: *if the derivative of a function vanishes identically, then it is a constant.* You can easily see that such a result is not true without the connectivity assumption on the boundary.

Definition 1.6.5 Let γ be a path defined on the interval $[a, b]$. It is called a *closed path* or a *loop* if its end points coincide, i.e., $\gamma(a) = \gamma(b)$. By a *point curve* is meant a curve with $Im(\gamma) = \{a\}$, i.e., the curve is given by a constant map.

We say γ is *simple* or *Jordan*[24] if it has no self-crossing, i.e., $\gamma(t_1) = \gamma(t_2) \Longleftrightarrow t_1 = t_2$ or $\{t_1, t_2\} = \{a, b\}$. By a *Jordan closed curve* or a *simple closed curve*, we mean a loop that is simple. Thus for a simple closed curve γ, it follows that $\gamma(a) = \gamma(b)$ and for pairs of distinct points $\{t_1, t_2\}$ other than $\{a, b\}$, we have, $\gamma(t_1) \neq \gamma(t_2)$.

An important fact about a simple closed curve is the so called:

Theorem 1.6.3 Jordan Curve Theorem:(JCT) *Every simple closed curve in $\mathbb{C}$ separates $\mathbb{C}$ into two regions, one bounded and another unbounded, and the curve is the common boundary of both of them.*

Remark 1.6.5 This theorem is geometrically quite self-evident, even to a 2-year old. Perform the following experiment and enjoy yourself. Equipment needed for this experiment are
(i) a piece of chalk (preferably a coloured one),
(ii) friendship with a two year old kid.
Draw a sufficiently large simple closed curve on the floor and ask the kid to sit inside it for two minutes while you pretend to look away but observe the out-come. Report your observation. (As a pre-caution, let the curve be really a simple one.)

However, a formal proof of JCT is not so easy[25]. In many expositions of Complex Analysis, JCT is used, and often, without proof. A proof for the case of a simple closed

[24]Carmile Jordan(1838-1922), a French mathematician.

[25]A proof due to a young Indian under graduate R. Munshi, is published in Resonance Vol **4** No. 9 (pp. 32-37) and No 11 (pp. 14-20) **1999**

curve which is piecewise smooth is somewhat easier and you will learn it in a Differential Topology course if at all you opt for it. You are welcome to try your hand to provide a proof in the case of a *polygonal Jordan loop.* As far as complex function theory is concerned, even this case is enough. (Compare Exercises 7.5.17 and 7.5.18.)

However, the word *boundary* used in the statement of JCT, may need some explanation. Since this will be useful later on, let us study this concept here.

Definition 1.6.6 Let $G \subseteq \mathbb{C}$. A point z is called a *limit point* of G if for every $\epsilon > 0$ we have, $(B_\epsilon(z) \setminus \{z\}) \cap G \neq \emptyset$. It is called an *interior point* of G if $B_\epsilon(z) \subset G$ for some $\epsilon > 0$. Observe that an interior point is certainly a limit point but the converse is not true. Points which are limit points but not interior points are called the *boundary points* of G. The set of boundary points of G is called the *boundary* of G and is denoted by δG. For nice subsets, this definition of boundary coincides with the intuitive notion of the boundary that we have. For instance, the boundary points of a disc are the points on the circle that bounds it. The set of interior points is called the *interior* of G and is denoted by *int* G. Observe that *int* G is always an open set and G together with all its limit points forms a closed set called the *closure of* G, denoted by $\bar{G}$. It is indeed the smallest closed set containing G. Of course, $\delta G = \bar{G} \setminus int\ G$ and hence is a closed set.

Exercise 1.6

1. Determine the number of connected components of the space $\mathbb{C} \setminus X$, where X is:
   (a) the real axis;
   (b) the parabola: $y = x^2$;
   (c) the line segment $[z_1, z_2]$ for any two points $z_1, z_2 \in \mathbb{C}$.

2. Show that any circle in $\mathbb{C}$ separates the complex plane into two components, one bounded and another unbounded, without of course using the JCT. Do it for any ellipse, a triangle, a rectangle, or any convex polygon.

3. Let Ω be a bounded convex open subset of $\mathbb{C}$. Show that $\mathbb{C} \setminus \partial\Omega$ has precisely two components, one bounded ($= \Omega$) and another unbounded.

4. Show that a path component of an open subspace A of $\mathbb{C}$ is open in $\mathbb{C}$. In general, they need be neither an open nor a closed subset of A.

5. If f is a continuous function on a connected set taking integer values then show that f is a constant function.

1.7 *Connectivity

There is a topological notion which is less intuitive but more fundamental than path-connectivity, viz., connectivity[26] , defined purely in terms of open subsets of a topological space. In this section, we shall study this notion, and relate it to the notion of path connectivity.

Many introductory expositions on real analysis include a proof of connectivity of an interval. Indeed, the axiom of existence of least upper bound for subsets of $\mathbb{R}$ which are bounded above (see (1.14), is **equivalent** to the statement that every interval in $\mathbb{R}$ is connected. Since we have not found any exposition containing this, we shall include a proof of this here.

In this section, we talk about 'spaces' to mean subsets of some Euclidean space. However, all results and concepts hold for any abstract topological space as well, with which, at this stage, you may not be familiar with. We begin with theorem 1.6.2 and make the concluding property in it into a definition.

Definition 1.7.1 Let X be a subset of $\mathbb{R}^n$ for some n. We say, that a subset $A \subset X$ is open in X if $A = X \cap U$ where U is open in $\mathbb{R}^n$. Of course A is closed in X if $X \setminus A$ is open in X. With this notion of open subsets of X, we shall call X a *space*.

Definition 1.7.2 A space $X \subseteq \mathbb{R}^n$ is called *connected* if the only nonempty subset of X which is both open and closed in X is X itself.

The following lemma gives you some equivalent formulation of connected spaces.

Lemma 1.7.1 Let Y be a subspace of $\mathbb{R}^n$ and X be a subset of Y. Then the following statements are equivalent:
(i) X is connected.
(ii) If X is contained in the disjoint union of two or more open subsets of Y then X is contained in one of them.
(iii) X is contained in the disjoint union of two closed subsets of Y then X is contained in one of them.

[26] It is possible to avoid this section completely for a long time. At least, in this book, the only places that we have actually used this is in the statement and the proof of theorem 7.4.2. But in the long run, the study of functions of one complex variable, will involve much deeper point-set-topology than the the basic results included in this section.

Proof: Easy. ♠

The following result comes very handy in proving that certain subsets are connected.

Theorem 1.7.1 *Let X be the union of a family of connected subsets C_α, where $\cap_\alpha C_\alpha$ is non empty. Then X is connected.*

Proof: Suppose A is a non empty, open as well as closed subset of X. Then $X = A\coprod(X \setminus A)$, where both A and $X \setminus A$ are closed. Since each C_α is connected it has to be contained either in A or in $X \setminus A$. Since $A \neq \emptyset$, $C_\beta \subset A$ for some β. Now for any other α, $\emptyset \neq C_\alpha \cap C_\beta \subset A \subset C_\alpha \cap A$ and hence $C_\alpha \subset A$. Therefore $X = A$. ♠

Remark 1.7.1

1. It follows from (iii) that the closure of a connected subset is connected. For if A is connected and $\bar{A} \subset F_1 \coprod F_2$, is in a disjoint union of two closed sets, then from (iii) A is contained in, say F_1. This then implies $\bar{A} \subset F_1$.

2. Maximal connected subsets are called connected components. Every space is the disjoint union of its connected components.

3. From (1) above, it follows that each connected component of a space is a closed subset. In general, components of a space need not be open subsets.

One of the main properties of connectivity is that it is preserved under continuous maps:

Theorem 1.7.2 *Let $f : X \to Y$ be a continuous map. If X is connected, then $f(X)$ is connected.*

Proof: If not, let $f(X) = U_1 \coprod U_2$ be the disjoint union of two open sets. Since f is continuous, it follows that $V_j = f^{-1}(U_j)$ are open in X. But it is easy to check that $X = V_1 \coprod V_2$. This means X is not connected. ♠.

Theorem 1.7.3 *A subset of $\mathbb{R}$ is connected iff it is an interval.*

Proof: Recall a subset A of $\mathbb{R}$ is called an interval if $a, b \in A$ and $a < x < b$ implies $x \in A$. So, if A is not an interval, then there exist $a, b \in A$ and $a < x < b$ such that $x \notin A$. We take $U_1 = (-\infty, x)$ and $U_2 = (x, \infty)$ and check that $A \subset U_1 \cup U_2$. Observe that both U_j are open in $\mathbb{R}$ and $U_1 \cap U_2 = \emptyset$. However, A is contained neither in U_1 nor in U_2. This means that A is not connected. Thus we have proved that if A is connected then it is an interval.

Conversely, suppose that A is an interval and we have two disjoint open sets U_1 and U_2 such that $A \subset U_1 \cup U_2$. If possible suppose $A \cap U_j \neq \emptyset, j = 1, 2$. Pick up $a_j \in A \cap U_j$. We may assume that $a_1 < a_2$. Put

$$Y = \{x \in A \cap U_1 \ : \ x < a_2\}.$$

Then Y is a non empty subset of $\mathbb{R}$ which is bounded above. Let y be the least upper bound for Y. (This is where we have used the least upper bound property; see section 2) Then clearly $a_1 \leq y \leq a_2$. Therefore $y \in A$. Suppose $y \in U_1$. This means there is an $0 < \epsilon < (a_2 - y)/2$ such that $[y, y + \epsilon) \subset U_1$. This implies $[y, y + \epsilon) \subset Y$. But then y cannot be the least upper bound of Y. Therefore $y \notin U_1$. Therefore $y \in U_2$. This means that for some $0 < \epsilon < y - a_1/2$, $(y - \epsilon, y] \subset U_2$. On the other hand, since y is the least upper bound, there exist $z \in (y - \epsilon, y) \in U_1$ which means $U_1 \cap U_2 \neq \emptyset$. This is absurd. Therefore one of the two sets $A \cap U_j$ is empty, which is the same as saying that A is contained in one of the two sets U_1 or U_2. This proves that A is connected. ♠

Remark 1.7.2 In terms of definition 1.7.2, theorem 1.6.2 can be stated as :*every path connected space is connected.* Combining theorems 1.7.2 and 1.7.3 we can obtain an alternate proof of this theorem. It is time you verified various properties stated in the previous section for connected spaces. In particular, we have:

Theorem 1.7.4 *A locally constant function on a connected space is a constant.*

Remark 1.7.3 Theorem 1.7.3 implies in particular, that $\mathbb{R}$ itself is connected. Indeed, we shall now show that the assumption $\mathbb{R}$ is connected implies that the least upper bound property (1.14) holds.

Let A be any non empty subset of $\mathbb{R}$ which is bounded above by say b. Assume that there is no least upper bound for A. The A is infinite. Let U be the set of all upper bounds for A. The $b \in U$ and $A \not\subset U$, U is a proper subset of $\mathbb{R}$. Our aim is to prove that U is both open and closed thereby getting a contradiction.

Let $x \in U$. Since x is not a least upper bound for A, there exists $y < x$ such that $(y, x] \cap A = \emptyset$. Since x is an upper bound for A we have $A \subset (-\infty, x]$. Therefore, $A \subset (-\infty, y]$. This means $(y, x] \subset U$. Since in any case all numbers bigger than x are also in U, it follows that $(y, \infty) \subset U$. This proves that U is open.

It remains to see that $\mathbb{R} \setminus U$ is also open. Let $x \notin U$. This means there is $a \in A$ such that $x < a$. Moreover $(-\infty, a) \cap U = \emptyset$. This shows that $\mathbb{R} \setminus U$ is also open. This completes the proof that connectivity of $\mathbb{R}$ implies the lub property (1.14).

Remark 1.7.4 Along the same line we can also prove that connectivity of intervals is equivalent to the intermediate value property. (See theorem 13.) Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Since $[a, b]$ is connected, it follows $f[a, b]$ is connected. Therefore it is an interval. (This proves theorem 13.) Conversely, suppose some interval A in $\mathbb{R}$ is not connected. Then $A \subset U_1 \coprod U_2$ a disjoint union of open sets and $A \cap U_j \neq \emptyset$. Choose $a_j \in A \cap U_j$ and assume that $a_1 < a_2$ for definiteness. Define a function f as follows: $f(x) = 1$ if $x \in U_1$, and $f(x) = 2$ if $x \in U_2$. Then f is continuous on $U_1 \coprod U_2$ and hence by restriction defines a continuous function on $[a_1, a_2] \subset A \subset U_1 \coprod U_2$. But $f[a_1, a_2] = \{1, 2\}$ which is not an interval. This shows that IVP is violated.

Finally, we shall include a technical result which becomes very handy while dealing with open connected subsets of Euclidean spaces.

Theorem 1.7.5 *If A is a connected open subset of a euclidean space. Then given any two points $a, b \in U$ there exist a path from a to b made up of finitely many line segments each of which is parallel to one of the axis. In particular A is path connected.*

Proof: Let us call a path as prescribed in the theorem a 'special' path. Let U be the subset of all points in A which can be joined to the point a by a special path in A. Clearly $a \in U$.

We shall show that U and $A \setminus U$ are both open. This will mean, by the connectivity of A that $U = A$. That will of course prove that $b \in U$.

Given $z \in \mathbb{R}^n$ and $\delta > 0$ let $Box_\delta(z)$ denote the cubical box with center z and each of its sides of length 2δ :

$$Box_\delta(z) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ |x_j - z_j| < \delta, 1 \leq j \leq n\}.$$

It is clear that any two points in a box as above can be joined by a special path in the box.

Let now $z \in U$. Fix a special path γ_1 in A from z_0 to z. Since A is open, there exists $\delta > 0$ such that the cubical the disc $B_\delta(z) \subseteq A$. Let γ_2 be a special path from z to w in the box where w is an arbitrary point of the box. Then $\gamma_1 \star \gamma_2$ is a special path in A from z_0 to w. Hence $w \in U$. This means that $Box_\delta(z) \subseteq U$. Therefore U is open.

On the other hand, suppose now that z is not a point of U. Then no point w of $Box_\delta(z)$ could have been joined to z_0 by a special path in A, for otherwise we could join z also to z_0, by first joining z to w inside $B_\delta(z)$ by a special path. This means $Box_\delta(z) \subseteq A \setminus U$. Hence $A \setminus U$ is also open. This completes the proof. ♠

Example 1.7.1 The Topologists Sine Curve: We end this section with the following example. Let

$$X = \left\{ \left(x, \sin\frac{1}{x} \right) \ : \ 0 < x \leq \pi \right\} \text{ and } Y = \{\imath y \ : \ -1 \leq y \leq 1\}.$$

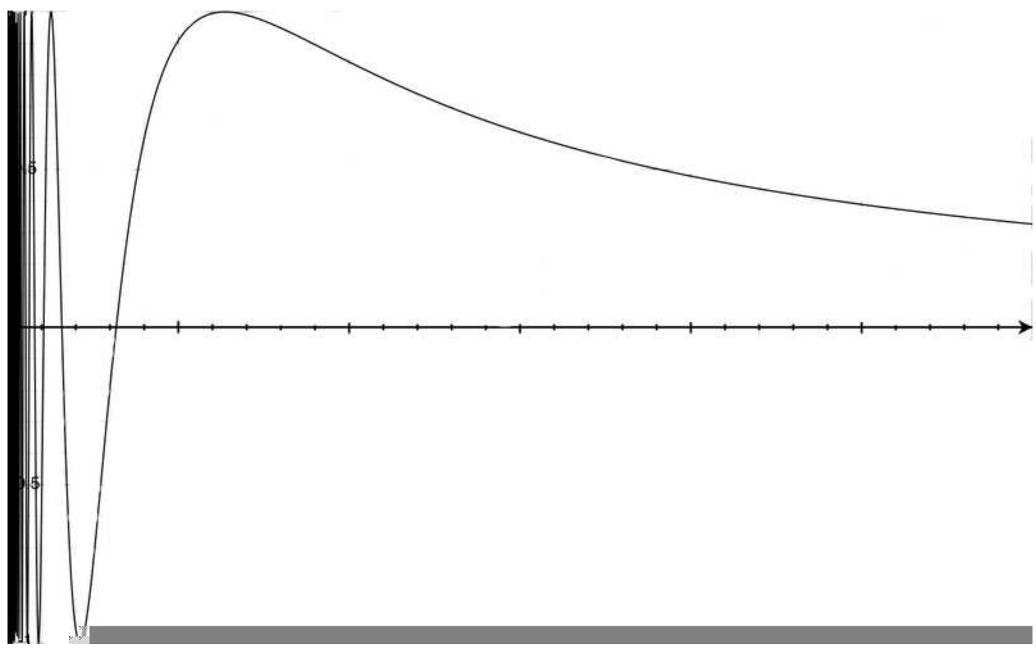


Fig. 9

Consider the subspace $Z = X \cup Y$ of $\mathbb{R}^2$. Both X and Y are path connected and in particular connected also. Observe that the closure of X is the whole space Z and hence it is connected. However, the space Z is not path connected. To see this suppose $\gamma : [0,1] \to Z$ is a path starting at $(0,0)$ and ending say at $(\pi, 0)$. Let t_0 be the least upper bound of all those t such that $\gamma(t) \in Y$. Since Y is a closed subset of Z, it follows that $\gamma(t_0) \in Y$. Put $U = B_{1/2}(\gamma(t_0)) \cap Z$ and observe that the projection $p_2(U)$ of U onto the real axis consists of a countable union of disjoint open intervals on the positive real axis and the point 0. On the other hand by continuity, there exists $\epsilon > 0$ such that $\gamma(t_0 - \epsilon, t_0 + \epsilon) \subset B_{1/2}(\gamma(t_0)) \cap Z$ and hence $p_2 \circ \gamma(t_0 - \epsilon, t_0 + \epsilon)$ is an interval contained in $p_2(U)$. Since 0 belongs to this interval, it follows that $p_2 \circ \gamma(t_0 - \epsilon, t_0 + \epsilon) = \{0\}$. This is the same as saying that $\gamma(t_0 - \epsilon, t_0 + \epsilon) \subset Y$ which contradicts the definition of t_0. It should be remarked that the space Z is not 'locally connected' at points of Y and this is essentially the reason how a connected space fails to be path connected.

Exercise 1.7 Show that product of two connected spaces is connected.

1.8 *The Fundamental Theorem of Algebra

As promised before, we shall give an elementary proof of Fundamental Theorem of Algebra (FTA) in this section.

Theorem 1.8.1 *Every non constant polynomial in one variable with coefficients in $\mathbb{C}$ has at least one root in $\mathbb{C}$.*

The proof uses only elementary Real Analysis which you have learnt so far. All proofs of FTA use Intermediate Value Theorem (IVP) implicitly or explicitly. We shall use it here explicitly. Apart from that, the only important result that we use is Weierstrass's theorem (1.5.5). The proofs of these results have been indicated in the form of exercises at the end of section 5. Those of you who have not seen these proofs yet may now read the solution.

We begin with:

Lemma 1.8.1 For every polynomial function $p : \mathbb{C} \to \mathbb{C}$, the function $|p| : \mathbb{C} \to \mathbb{R}$ attains its infimum.

Proof: Given a polynomial p, we have to show that there exists $z_0 \in \mathbb{C}$ such that $|p(z_0)| \leq |p(z)|$ for all $z \in \mathbb{C}$.

In exercise 11 in section 1.1, we have seen that $p(z) \longrightarrow \infty$ as $z \longrightarrow \infty$. This means that there exists $R > 0$ such that $|p(z)| > |p(0)|$ for all $|z| > R$. It follows that

$$\text{Inf } \{|p(z)| \ : \ z \in \mathbb{C}\} = \text{Inf } \{|p(z)| \ : \ |z| \leq R\} \leq |p(0)|.$$

But the disc $\{z \ : \ |z| \leq R\}$ is closed and bounded. Since the function $z \mapsto |p(z)|$ is continuous, it attains its infimum on this disc. This completes the proof of the lemma. ♠

Slowly but surely, now an idea of the proof of FTA emerges: Observe that FTA is true iff the infimum z_0 obtained in the above lemma is a zero of p, i.e., $p(z_0) = 0$.

Therefore in order to complete a proof of FTA, it is enough to assume that $p(z_0) \neq 0$ and arrive at a contradiction. (This idea is essentially due to Argand.)

Consider the polynomial $q(z) = p(z + z_0)$. Both the polynomials, p, q have the same value set and hence minimum of $|q(z)|$ is equal to minimum of $|p(z)|$ which is equal to $|p(z_0)| = |q(0)|$.

We shall assume that $q(0) \neq 0$ and arrive a contradiction.

Write $q(z) = q(0)\phi(z)$. Then $|q(0)|$ is the minimum of $|q(z)|$ iff 1 is the minimum of $|\phi(z)|$. Put

$$\phi(z) = 1 + wz^k + z^{k+1}f(z)$$

with $w \neq 0$ is some complex number, $k \geq 1$ and $f(z)$ some polynomial.

It is enough to prove that

Lemma 1.8.2 Argand's Inequality For any polynomial f, positive integer k, and any $w \in \mathbb{C} \setminus \{0\}$,

$$\text{Min}\{|1 + wz^k + z^{k+1}f(z)| \ : \ z \in \mathbb{C}\} < 1. \tag{1.49}$$

Choose $r > 0$ such that $r^k = |w|$ (IVP)(see Exercise 1.5.13). Now replace z by z/r^k in (1.49). Thus, we may assume $|w| = 1$ in (1.49).

At this stage, Argand's proof uses de Moivre's theorem, viz., for every complex number α and every positive integer k, the equation $z^k = w$ has a solution. For its simplicity, we present this proof first:

Choose λ such that $\lambda^k = -w^{-1}$. Replace z by λz in (1.49) to reduce it to proving

$$\text{Min}\{|1 - z^k + z^{k+1}g(z)| \ : \ z \in \mathbb{C}\} < 1. \tag{1.50}$$

Now restrict z to positive real numbers, $z = t > 0$. Since $g(t)$ is a polynomial, $tg(t) \to 0$ as $t \to 0$. So there exists $0 < t < 1$ for which $|tg(t)| < 1/2$. But then

$$|1 - t^k + t^{k+1}g(t)| < |1 - t^k| + \frac{t^k}{2} = 1 - t^k + \frac{t^k}{2} < 1$$

thereby completing the proof of (1.49).

Why do we want to avoid using de Moivre's Theorem? The answer is that it depends heavily upon the intuitive concept of the angle which needs to be established rigorously. (It should also be noted that during Argand's time, one could not expect a rigorous proof of lemma 1.8.1, which Argand simply assumed.[27])

[27]For more learned comments, see R. Remmert's article on 'Fundamental Theorem of Algebra' in [Ebb].

Instead, we now follow an idea of Littlewood[28] which is coded in the following two lemmas:

Lemma 1.8.3 Given any complex number w of modulus 1, one of the four numbers $\pm w, \pm \imath w$ has its real part less than $-1/2$.

Proof: [This is seen easily as illustrated in the Fig. 9. The four shaded regions which cover the whole of the boundary are got by rotating the region $\Re(z) < -1/2$. However, it is important to note that the following proof is completely independent of the picture.] Since $|w| = 1$, either $|\Re(w)|$ or $|\Im(w)|$ has to be bigger than $1/2$. In the former case, one of $\pm w$ will have the required property. In the latter case, one of $\pm \imath w$ will do. ♠

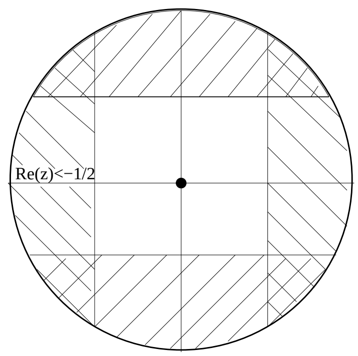


Fig. 9

Lemma 1.8.4 For any integer $n \geq 1$, the four equations

$$z^n = \pm 1; \quad z^n = \pm \imath; \tag{1.51}$$

have all solutions in $\mathbb{C}$.

Proof: Write $n = 2^k m$, where $m = 4l + 1$ or $4l + 3$. For $k \geq 0$, since we can take successive square-roots (see Example 1.2.1), let $\alpha_k, \beta_k, \gamma_k$ be such that

$$\alpha_k^{2^k} = -1, \quad \beta_k^{2^k} = \imath, \quad \gamma_k^{2^k} = -\imath.$$

(For $k = 0$, this just means $\alpha_0 = -1; \beta_0 = \imath, \gamma_0 = -\imath$.)

Now let us solve the four equations (1.51) one by one:

(a) For $z^n = 1$, take $z = 1$.

(b) For $z^n = -1$, take $z = \alpha_k$. Then $(\alpha_k)^n = (-1)^m = -1$.

(c) For $z^n = \imath$, there are two cases: If $m = 4l + 1$, then take $z = \beta_k$ so that $(\beta_k)^n =$

[28]??

$(\imath)^m = \imath$. If $m = 4l + 3$ then take $z = \gamma_k$ so that $(\gamma_k)^n = (-\imath)^m = (-\imath)^3 = \imath$.
(d) This case follows easily from (b) and (c) : Choose z_1, z_2 such that $z_1^n = -1$ and $z_2^n = \imath$. Then $(z_1 z_2)^n = -\imath$. ♠

[At this stage, the proof given in literature first establishes de Moivre's theorem and then follows the arguments as in Argand's proof. Here, we shall directly derive Argand's inequality.]

Returning to the proof of lemma 1.8.2, choose $\tau = \pm 1, \pm\imath$ so that $\Re(\tau w) < -\dfrac{1}{2}$ (Lemma 1.8.3). Choose $\alpha \in \mathbb{C}$ such that $\alpha^k = \tau$ (Lemma 1.8.4).

Now, replace z by αz, so that we may assume that $w = a + \imath b$, where $a \leq -1/2$ and $a^2 + b^2 = 1$.

Since f is continuous, it follows that $tf(t) \to 0$ as $t \to 0$. Restricting to just the real values of t, we can choose $0 < \delta < 1$ such that $|tf(t)| < \frac{1}{3}$ for all $0 < t < \delta$. For such a choice of t, we have

$$|1 + wt^k + t^{k+1}f(t)| \leq |1 + wt^k| + \frac{t^k}{3} = [(1 + at^k)^2 + b^2t^{2k}]^{1/2} + \frac{t^k}{3}.$$

We want to choose $0 < t < \delta$ such that this quantity is less than 1. For $a^2 + b^2 = 1$ and $t > 0$ we have

$$\begin{array}{lll}
& [(1 + at^k)^2 + b^2t^{2k}]^{1/2} + \dfrac{t^k}{3} < 1 & \text{iff} \quad [(1 + at^k)^2 + b^2t^{2k}]^{1/2} < 1 - \dfrac{t^k}{3} \\
\text{iff} & (1 + at^k)^2 + b^2t^{2k} < \left(1 - \dfrac{t^k}{3}\right)^2 & = \quad 1 - \dfrac{2t^k}{3} + \dfrac{t^{2k}}{9} \\
\text{if} & 1 + 2at^k + t^{2k} < 1 - \dfrac{2t^k}{3} + \dfrac{t^{2k}}{9} & \quad \text{since } b^2 \leq 1 \\
\text{iff} & \dfrac{8}{9}t^k < -\left(2a + \dfrac{2}{3}\right), \quad t > 0. &
\end{array}$$

This last condition can be fulfilled by choosing $t > 0$ such that $t^k < 3/8$, for then,

$$\frac{8}{9}t^k < \frac{1}{3} < -\left(2a + \frac{2}{3}\right).$$

Thus, for any $t > 0$ which is such that $t^k < \min\{3/8, \delta\}$ (IVP again), we have

$$|1 + wt^k + t^{k+1}f(t)| < 1.$$

This completes the proof of the lemma 1.8.2 and thereby that of FTA. ♠

1.9 Miscellaneous Exercises to Ch. 1

1. Use de Moivre's theorem to prove:
   (*a*) $\cos 2\theta = \cos^2\theta - \sin^2\theta$;
   (*b*) $\sin 2\theta = 2\sin\theta\cos\theta$;
   (*c*) $\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta$;
   (*d*) $\sin 3\theta = 3\cos^2\theta\sin\theta - \sin^3\theta$.

2. For every positive integer n, put $\alpha_n = \sum_{k\geq 0}(-1)^k\binom{n}{2k}$; $\beta_n = \sum_{k\geq 0}(-1)^k\binom{n}{2k+1}$. Prove that

$$(\alpha_n, \beta_n) = \begin{cases} 2^{n/2}(1,0), & n \equiv 0(\text{mod}8); \\ 2^{(n-1)/2}(1,1), & n \equiv 1(\text{mod}8); \\ 2^{n/2}(0,1), & n \equiv 2(\text{mod}8); \\ 2^{(n-1)/2}(-1,1), & n \equiv 3(\text{mod}8); \\ 2^{n/2}(-1,0), & n \equiv 4(\text{mod}8); \\ 2^{(n-1)/2}(-1,-1), & n \equiv 5(\text{mod}8); \\ 2^{n/2}(0,-1), & n \equiv 6(\text{mod}8); \\ 2^{(n-1)/2}(1,-1), & n \equiv 7(\text{mod}8). \end{cases}$$

   Also prove that $\alpha_n^2 + \beta_n^2 = 2^n$.

3. Prove the following identities

   (i) $1 + \cos\theta + \cos 2\theta + \cdots + \cos n\theta = \dfrac{1}{2} + \dfrac{\sin(\frac{2n+1}{2})\theta}{2\sin\frac{\theta}{2}} \quad (0 < \theta < 2\pi)$

   (ii) $\sin\theta + \sin 2\theta + \cdots + \sin n\theta = \dfrac{1}{2}\cos\dfrac{\theta}{2} - \dfrac{\cos(\frac{2n+1}{2})\theta}{2\sin\frac{\theta}{2}} \quad (0 < \theta < 2\pi)$

   (iii) $\sin\theta + \sin 3\theta + \cdots + \sin(2n-1)\theta = \dfrac{\sin^2 n\theta}{\sin\theta}, \quad (0 < \theta < \pi)$.

   (iv) $\cos\theta + \cos 3\theta + \cdots + \cos(2n-1)\theta = \dfrac{\sin 2n\theta}{2\sin\theta} \quad (0 < \theta < \pi)$.

4. Let $\zeta = E(2\pi/n)$. Show that $(1-\zeta)(1-\zeta^2)\cdots(1-\zeta^{n-1}) = n$. Deduce

$$\sin\frac{\pi}{n}\sin\frac{2\pi}{n}\cdots\sin\frac{(n-1)\pi}{n} = \frac{n}{2^{n-1}}.$$

5. Let $P_1, P_2, \ldots, P_n$ be the vertices of a regular polygon inscribed in a circle of radius r. Determine the product of distances of $P_2, \ldots, P_n$ from P_1.

6. Describe geometrically the regions in the plane defined by:
   (a) $|3z-2|<1$; (b) $\Im z \geq 2$; (c) $0<\arg z<\pi/4$;
   (d) $|z-1|+|z+1|\leq 4$; (e) $\Re(z^2)+\Im(z^2)=0$; (f) $|z-\rho|=\rho+\Re z,\ \rho>0$.

7. (*Circles of Appolonius*) Determine the locus of the points given by
$$\left|\frac{1-z}{1+z}\right|=r$$
for any $r\geq 0$. What happens for $r=0,1,\infty$?

8. Using complex numbers show that the line joining the mid-points of two sides of a triangle is parallel to the third side and is of half its length. Also show that the midpoints of the sides of any quadrilateral form the vertices of a parallelogram.

9. For any two distinct complex numbers z, w, let us denote by $[z,w]$ the line segment from z to w. Show that two line segments $[z_1,w_1],[z_2,w_2]$ in the plane are parallel iff $(z_1-w_1)(\overline{z_2-w_2})$ is a real number. In particular, show that three points z_1,z_2,z_3 are collinear iff $(z_1-z_2)(\overline{z_1-z_3})$ is a real number.

10. Show that three points in the plane representing z_1,z_2,z_3 are collinear iff
$$\begin{vmatrix} 1 & z_1 & \bar{z}_1 \\ 1 & z_2 & \bar{z}_2 \\ 1 & z_3 & \bar{z}_3 \end{vmatrix}=0.$$

11. Show that the area of the triangle formed by three distinct points z_1,z_2,z_3 is given by
$$\frac{1}{2}\left|(z_1-z_3)^2\Im\left(\frac{z_1-z_2}{z_1-z_3}\right)\right|.$$
Deduce that the three points are collinear iff $\frac{z_1-z_2}{z_1-z_3}$ is real.

12. Let a,b be any two non zero complex numbers. Show that there exists $z\in\mathbb{C}$ such that $|z+a|+|z-a|=2|b|$ iff $|a|\leq|b|$.

13. Let z_1,z_2,z_3 and w_1,w_2,w_3 form the vertices of two similar triangles the labeling being taken in the counter-clockwise sense in both cases. Then show that
$$\begin{vmatrix} z_1 & w_1 & 1 \\ z_2 & w_2 & 1 \\ z_3 & w_3 & 1 \end{vmatrix}=0.$$
[Hint: First reduce the problem to the case when $z_1=w_1=0$. Next, reduce it to the case when $z_2=1$.]

14. Show that a cyclic quadrilateral is a rectangle if its centroid coincides with the center of the circle in which it is inscribed.

15. Take an arbitrary quadrilateral and raise a square exterior to the quadrilateral, on each of its side. Join the centers of the squares on the opposite sides to obtain two line segments. Use complex numbers to show that these line segments are equal in length and perpendicular to each other.

16. (**Nepolian's Theorem**) Take any triangle and raise equilateral triangles on each side of it, external to the triangle. Show that the centres of these triangles form an equilateral triangle.

17. Suppose $w_1, w_2, w_3 \in \mathbb{C}\backslash\{0\}$ represent three non-collinear points and $w_1+w_2+w_3 = 0$. If a_1, a_2, a_3 are real numbers such that $a_1w_1 + a_2w_2 + a_3w_3 = 0$ then show that $a_1 = a_2 = a_3$. (This is equivalent to say that if three planar forces acting on a point are keeping it in equilibrium then by scaling all the three forces by the same factor only, the point will be still in equilibrium.)

18. Let $w_1, w_2, w_3 \in \mathbb{C} \setminus \{0\}$ be such that $w_1 + w_2 + w_3 = 0$. Prove that the following statements are equivalent.
    (i) $w_1^2 + w_2^2 + w_3^2 = 0$.
    (ii) $w_1w_2 + w_2w_3 + w_3w_1 = 0$.
    (iii) $\dfrac{1}{w_1} + \dfrac{1}{w_2} + \dfrac{1}{w_3} = 0$.
    (iv) $|w_1| = |w_2| = |w_3|$.
    (v) $0, w_1, w_1 + w_2$ form the vertices of an equilateral triangle.
    (vi) w_1, w_2, w_3 form the vertices of an equilateral triangle.

19. Prove that three distinct points z_1, z_2, z_3 in the plane form the vertices of an equilateral triangle iff $z_1^2 + z_2^2 + z_3^2 = z_1z_2 + z_2z_3 + z_3z_1$. Deduce that if w_1, w_2, w_3 are points dividing the three sides of the triangle $\Delta(z_1, z_2, z_3)$ in the same ratio, then the triangle $\Delta(w_1, w_2, w_3)$ is equilateral iff the triangle $\Delta(z_1, z_2, z_3)$ is so.

20. Given a point $w \in \mathbb{C}$, find the expression for the foot of the perpendicular from w onto the line $L : \alpha z + \overline{\alpha z} = k$. Find the expression for the reflection of w in this line.

21. Let R_1, R_2, R_3 be reflections in sides of a genuine triangle. Show that their composite taken in any order is a glide-reflection.

22. Show that four points in the plane no three of which are collinear, representing z_1, z_2, z_3, z_4 are con-cyclic iff

$$\begin{vmatrix} 1 & z_1 & \bar{z}_1 & z_1\bar{z}_1 \\ 1 & z_2 & \bar{z}_2 & z_2\bar{z}_2 \\ 1 & z_3 & \bar{z}_3 & z_3\bar{z}_3 \\ 1 & z_4 & \bar{z}_4 & z_4\bar{z}_4 \end{vmatrix} = 0.$$

23. Show that if $z_1, z_2, \ldots, z_n$ all lie strictly on one side of a line through the origin, then $z_1^{-1}, \ldots, z_n^{-1}$ also lie on one side of a line passing through the origin; moreover prove that $\sum_j z_j \neq 0$ and $\sum_j \frac{1}{z_j} \neq 0$.

24. Given two distinct points $a, b \in \mathbb{C}$, show that the equations

$$\arg \frac{z-a}{z-b} = k$$

represent the family of circles passing through the two points.

25. Let z_1, z_2, z_3 be any three distinct points on the circle $|z| = r$. Suppose 0 and z_3 lie on the same side of the chord $[z_1, z_2]$. Show that $\arg \frac{z_2 - z_3}{z_1 - z_3} = \frac{1}{2} \arg \frac{z_2}{z_1}$.

26. For any four distinct complex numbers $z_j, j = 1, 2, 3, 4$ we define the their cross ratio by

$$[z_1, z_2, z_3, z_4] = \left(\frac{z_1 - z_3}{z_2 - z_3}\right)\left(\frac{z_2 - z_4}{z_1 - z_4}\right).$$

$z_j, j = 1, 2, 3, 4$ lie on a straight line or a circle iff $[z_1, z_2, z_3, z_4]$ is a real number. From this obtain the equation of a circle passing through three non-collinear points z_1, z_2, z_3.

27. Find the center of the mass of point-masses λ_j situated at three distinct points z_j, $i = 1, 2, 3$ respectively. Allowing λ_j to have negative values also, determine when the center of mass lies inside the triangle $\Delta(z_1, z_2, z_3)$. Generalize this to n points.

28. Let $\omega \neq 1$ denote a cube root of unity. Mark on the plane, all the point $m + n\omega$ where m, n range over all integers. Show that by joining these points appropriately, the entire plane can be divided into non overlapping equilateral triangles.

29. Let $\omega \neq 1$ denote a cube root of unity. Find the minimum non zero value of $|m\omega + n\omega^2|$ where m, n range over $\mathbb{Z}$.

30. Show that the set $\mathbb{S}^1$ of all complex numbers of modulus 1 forms a multiplicative abelian (i.e., commutative) group. Also, show that for each natural number n, the set of n^{th} roots of unity forms a subgroup of $\mathbb{S}^1$.

31. Let $z \in \mathbb{C}$ be such that $|z| = 1$. Suppose there are integers $k_1 < k_2$ such that $1 + z^{k_1} + z^{k_2} = 0$. Show that z is a n^{th} root of unity.

32. Let $z \in \mathbb{C}$ be such that $|z| = 1$. Suppose there are integers $k_1 < k_2 < k_3$ such that $1 + z^{k_1} + z^{k_2} + z^{k_3} = 0$. Show that $z^n = 1$ for some integer n. In general, given $z \in \mathbb{C}$ and integers $1 < k_1 < k_2 \cdots < k_r$ where $r \geq 4$, such that $1 + z^{k_1} + \cdots + z^{k_r} = 0$, one can ask: " is $z^n = 1$ for some integer n?" This question was posed by Prof. M.G. Nadkarni[29] sometimes in 1993 to the author and answered in the negative by Praneshachar[30] in 1997, via the following example:

33.* Consider the polynomial

$$P(z) = 1 + z + z^3 + z^5 + z^6.$$

Put $z + \frac{1}{z} = r$ in $\dfrac{P(z)}{z^3}$ to obtain a cubic polynomial $f(r)$. Show that
(i) r is a root of f implies that z is root of P.
(ii) $f(r)$ has a root r_0 in the interval $(-1, 1)$.
(iii) With $\theta = \cos^{-1}(r_0/2)$ and $z_0 = E(\theta)$ we have, $|z_0| = 1$ and $P(z_0) = 0$.
(iv) Show that P does not have any root which is a root of unity.

34. Let C be a circle around 0 so that $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$ has no zeros outside C. Define $f(z) = 1/p(z)$ outside C. Show that f is bounded outside C.

35. Let p be a non constant polynomial. Given $r > 0$ show that there exists $R > 0$ such that for any $|w| > R$, there exists z such that $p(z) = w$ and $|z| > r$.

36. Show that the amount of rotation in a given rigid motion is independent of the choice of origin or axis of reference used in expressing the rigid motion.

37. Given $w_1 \neq w_2 \in \mathbb{C}$, the reflection in the perpendicular divider of the segment $[w_1, w_2]$ maps w_1 to w_2 and vice versa. What are all the rigid motions which map w_1 to w_2?

[29]Department of Mathematics, University of Bombay
[30]Department of Mathematics, I. I. Sc. Bangalore

38. Given two sets of distinct points $\{z_1, z_2\}$ and $\{w_1, w_2\}$ such that $d(z_1, z_2) = d(w_1, w_2)$, show that there is a rigid motion which is the composite of at most two reflections which maps z_j to $w_j, i = 1, 2$.

39. Given two sets of distinct points $\{z_1, z_2, z_3\}$ and $\{w_1, w_2, w_3\}$ such that $d(z_j, z_k) = d(w_j, w_k), j, k = 1, 2, 3$, show that there exists a rigid motion which maps z_j to $w_j, j = 1, 2, 3$ which is the composite of at most three reflections.

40. Deduce, from the previous exercise, that every rigid motion of the plane is a composite of at most three reflections. Also, deduce this directly from theorem 1.3.1. Is there any kind of uniqueness in this decomposition?

41. Let p, q be positive real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$.
(a) Find the points of maxima-minima of the function $\phi(x) = \frac{1}{q} + \frac{1}{p}x - x^{1/p}$ on $[0, \infty)$. Deduce that $x^{1/p} \leq \frac{1}{q} + \frac{1}{p}x$, for all $x \geq 0$. (Equality hold iff $x = 1$.)
(b) Prove that for $u, v \geq 0$,
$$uv \leq \frac{u^p}{p} + \frac{v^q}{q}.$$
Show that equality holds iff $u^p = v^q$.
(c) Let $a_j, b_j \geq 0, j = 1, 2, \ldots, n$. Suppose $\sum_j a_j^p = 1 = \sum_j b_j^q$. Show that $\sum_j a_j b_j \leq 1$.
(d) Prove **Hölder's Inequality:** For any complex numbers $z_j, w_j, j = 1, 2, \ldots, n$ we have
$$\left|\sum_{j=1}^n z_j w_j\right| \leq \left(\sum_{j=1}^n |z_j|^p\right)^{1/p} \left(\sum_{j=1}^n |w_j|^q\right)^{1/q}.$$
(e) Deduce **Schwarz's inequality:**
$$\left|\sum_{j=1}^n z_j w_j\right| \leq \left(\sum_{j=1}^n |z_j|^2\right)^{1/2} \left(\sum_{j=1}^n |w_j|^2\right)^{1/2}.$$
(f) Deduce **Minkowski Inequality:**
$$\left(\sum_{j=1}^n |z_j + w_j|^p\right)^{1/p} \leq \left(\sum_{j=1}^n |z_j|^p\right)^{1/p} \left(\sum_{j=1}^n |w_j|^p\right)^{1/p}.$$

Chapter 2

Complex Differentiability

2.1 Definition and Basic Properties

Recall that for a real valued function f defined in an open interval, and a point x_0 in the interval, we say f is differentiable at x_0 if the limit of the difference quotient

$$\lim_{h \longrightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. Moreover, this limit is then called the derivative of f at x_0 and is denoted by $\frac{df}{dx}$ or by $f'(x_0)$.

In order to talk about differentiability of a function f of a real variable at a point x, observe that the map should be defined in an interval around x. Similarly, in case of a function f of a complex variable, we shall need that the function is defined in a disc of radius $r > 0$ around the point under consideration. Just to avoid the necessity of mentioning this condition every time, we briefly recall the concept of an open set here. (For more details, see section 1.5.)

Definition 2.1.1 A subset $U \subset \mathbb{C}$ is called an open set if for each point $z \in U$, we have $r > 0$ such that the open-ball $B_r(z) \subset U$ where,

$$B_r(z) = \{w \in \mathbb{C} \ : \ |w - z| < r\}.$$

Let us now consider a complex-valued function f defined in an open subset of $\mathbb{C}$ and define the concept of differentiation with respect to the complex variable. With no valid justification or motivation to do otherwise, we opt for a similar definition of differentiability of f in this case also as in the case of a real valued function of a real variable, as a limit of 'difference quotients'. All that we need is that these 'difference quotients' make sense.

Definition 2.1.2 Let $z_0 \in U$, where U is an open subset of $\mathbb{C}$. Let $f : U \longrightarrow \mathbb{C}$ be a map. Then f is said to be *complex differentiable* (written $\mathbb{C}-$*differentiable*) at z_0 if the limit on the right hand side of (2.1) exists, and in that case we call this limit, the derivative of f at z_0 :

$$\boxed{\boxed{\frac{df}{dz}(z_0) := \lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h}.}} \tag{2.1}$$

We also use the notation $f'(z_0)$ for this limit and call it *Cauchy derivative* of f at z_0.

If f is $\mathbb{C}-$differentiable at each $z \in U$ then the map $z \mapsto f'(z)$ is called the derivative of f on U and is denoted by f'.

Example 2.1.1 Let us work out the derivative of the function $f(z) = z^n$, where n is an integer, directly from the definition. Of course, for $n = 0$, the function is a constant and hence, it is easily seen that it is differentiable everywhere and the derivative vanishes identically. Consider the case when n is a positive integer. Then by binomial expansion, we have,

$$f(z+h) - f(z) = h\left(\binom{n}{1}z^{n-1} + \binom{n}{2}hz^{n-2} + \cdots + h^{n-1}\right).$$

Therefore, we have,

$$\lim_{h\longrightarrow 0} \frac{f(z+h)-f(z)}{h} = nz^{n-1}.$$

This is valid for all values of z. Hence f is differentiable in the whole plane and its derivative is given by $f'(z) = nz^{n-1}$. Next, consider the case when n is a negative integer. We see that the function is not defined at the point $z = 0$. Hence we consider only points $z \neq 0$. Writing $n = -m$ and $f(z+h) - f(z) = \dfrac{z^m - (z+h)^m}{(z+h)^m z^m}$ and applying binomial expansion for the numerator as above, we again see that

$$f'(z) = -\frac{m}{z^{m+1}} = nz^{n-1}, \; z \neq 0. \tag{2.2}$$

Remark 2.1.1 As in the case of calculus of 1-real variable, the Cauchy derivative has all the standard properties:

(i) The sum $f_1 + f_2$ of two $\mathbb{C}-$differentiable functions f_1, f_2 is $\mathbb{C}-$differentiable and

$$(f_1 + f_2)'(z) = f_1'(z) + f_2'(z).$$

(ii) the scalar multiple of a complex differentiable function f is complex differentiable, and

$$(\alpha f)'(z) = \alpha f'(z).$$

(iii) The product of two complex differentiable functions f, g is again complex differentiable and we have the product rule:

$$(fg)'(z) = f'(z)g(z) + f(z)g'(z). \tag{2.3}$$

Further if $g(z) \neq 0$ then we have the quotient rule:

$$\left(\frac{f}{g}\right)'(z) = \frac{f'(z)g(z) - f(z)g'(z)}{g^2(z)}. \tag{2.4}$$

The proof of the following theorem is exactly the same as the proof of the corresponding result for real valued function of a real variable.

Theorem 2.1.1 (**The Increment Theorem:**) *Let* $f : A \longrightarrow \mathbb{C}$, $z_0 \in A$, $r > 0$ *such that* $B_r(z_0) \subset A$. *Then* f *is complex differentiable at* z_0 *iff* $\exists\ \alpha \in \mathbb{C}$, *and a function* $\phi : B_s(0) \setminus \{0\} \longrightarrow \mathbb{C}$, $(0 < s \leq r)$ *such that for all* $h \in B_s(0) \setminus \{0\}$, *we have,*

$$\boxed{\boxed{f(z_0 + h) - f(z_0) = h\alpha + h\phi(h);\ \ \lim_{h\to 0} \phi(h) = 0.}} \tag{2.5}$$

Proof: For the given f, we simply take

$$\psi(h) := \frac{f(z_0 + h) - f(z_0)}{h}, \quad h \in B_r(0) \setminus \{0\}.$$

Then f is complex differentiable at z_0 iff $\lim_{h\longrightarrow 0} \psi(h)$ exists. In that case, we simply put α equal to this limit and take $\phi(h) = \psi(h) - \alpha$ and observe that $\phi(h) \longrightarrow 0$ iff $\psi(h) \longrightarrow \alpha$. On the other hand, if there is such a function ϕ and a constant α then clearly, $\lim_{h\longrightarrow 0} \psi(h) = \alpha$, and so, f is complex differentiable at z_0 and $f'(z_0) = \alpha$. ♠

Remark 2.1.2 We may assume that the error function ϕ in (2.5) is defined on the whole of $B_r(0)$, its value at 0 being completely irrelevant for us. The increment theorem enables one to deal with many tricky situations while dealing with differentiability. From (2.5) one can quickly deduce that a function which is differentiable at a point is continuous at that point. As a further illustration we shall derive the chain rule for differentiation.

Theorem 2.1.2 Chain Rule : *Let $f : A \longrightarrow \mathbb{C}$, $g : B \longrightarrow \mathbb{C}$, $f(A) \subset B$ and $z_0 \in A$. Suppose that $f'(z_0)$ and $g'(f(z_0))$ exist. Then $(g \circ f)'(z_0)$ exists and $(g \circ f)'(z_0) = g'(f(z_0))f'(z_0)$.*

Proof: Apply increment theorem to f at z_0 and to g at $f(z_0)$ to obtain:

$$\left.\begin{array}{rcll} f(z_0+h) - f(z_0) & = & hf'(z_0) + h\eta(h); & \eta(h) \to 0 \text{ as } h \to 0; \\ g(f(z_0)+k) - g(f(z_0)) & = & kg'(f(z_0)) + k\zeta(k) \ ; & \zeta(k) \to 0 \text{ as } k \to 0 \end{array}\right\} \tag{2.6}$$

Let η, ζ be defined over $B_s(0), B_r(0)$ respectively. By continuity of f at z_0, it follows that if s is chosen sufficiently small then for all $h \in B_s(0)$, we have $f(z_0+h) - f(z_0) \in B_r(0)$. Hence we can put $k = f(z_0+h) - f(z_0)$, in (2.6) to obtain,

$$\begin{array}{rcl} g(f(z_0+h)) - g(f(z_0)) & = & [hf'(z_0) + h\eta(h)][g'(f(z_0)) + \zeta(k)] \\ & = & hg'(f(z_0))f'(z_0) + h\xi(h), \end{array}$$

where, $\xi(h) = \eta(h)g'(f(z_0)) + (\eta(h) + f'(z_0))\zeta(f(z_0+h) - f(z_0))$. Observe that as $h \to 0$, we have, $k = f(z_0+h) - f(z_0) \to 0$ and $\zeta(k) \to 0$. Hence $\xi(h) \to 0$, as $h \to 0$. Thus by the increment theorem again, $(g \circ f)'(z_0)$ exists and is equal to $g'(f(z_0))f'(z_0)$ as desired.

Remark 2.1.3 So far, we have considered derivatives of functions at a point $z \in A$ only when the function is defined in a neighborhood of z. We can try to relax this condition as follows: Thus, if $B \subset \mathbb{C}$ and $f : B \longrightarrow \mathbb{C}$, we say f is complex differentiable on B if f extends to a complex differentiable map $\hat{f} : A \longrightarrow \mathbb{C}$ where A is an open subset containing B. However, we can no longer attach a unique derivative to f at points of B in general. With some more suitable geometric assumptions on B, this can be made possible. For instance if B is a closed disc or a closed rectangle, and if $f : B \longrightarrow \mathbb{C}$ is differentiable on B, then even at the boundary points of B, the derivative of f is unique. This follows since, once the limit of the difference quotient exists its value is determined by the values of f itself rather than those of the extensions of f.

Exercise 2.1

1. Write down detailed proofs of each of the claims made in remark 2.1.1.

2. Check for differentiability of the following functions directly from the definition at $z = 0$ and explain what goes wrong in case it is going wrong:
   (a) $z^2 + z + 1$; (b) $z^{1/2}$; (c) $\bar{z}$.

3. Use chain rule, product rule and (2.2) to obtain the quotient rule.

4. Use quotient rule, chain rule etc. to find the derivatives of
   $(a)\ \dfrac{z}{1+z}$; $(b)\ \dfrac{z^{1/2}}{z^2+z+1}$, wherever they exist.

5. Let $f : U \longrightarrow \mathbb{C}$ be a complex differentiable function, where U is a convex open subset of $\mathbb{C}$. Observe that if $g : J \to U$ is a real diffeentiable function on an open interval J, then the chain rule is valid for the composite function $f \circ g$. Use this to show that if $f'(z) \equiv 0$ then f is a constant function. Generalize this to the case when U is any connected open set. [Hint: Use theorem 1.7.5.]

2.2 Polynomials and Rational Functions

In this section, let us study some simple examples of complex differentiable functions:

(a) The polynomial functions: As such, the easiest functions to deal with are the constant functions. These are the first examples of $\mathbb{C}$-differentiable functions. The identity function $z \mapsto z$, merely denoted by z, is also $\mathbb{C}$-differentiable. Since product and sum of any two $\mathbb{C}$-differentiable function is again complex differentiable, it follows that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n,$$

is again $\mathbb{C}$-differentiable, where $a_j \in \mathbb{C}$. Such functions are called polynomials functions. If $a_n \neq 0$ then we say the degree of this polynomial is n. Moreover, the derivative of p with respect to z is given by

$$p'(z) = a_1 + 2a_2 z + \cdots n a_n z^{n-1}.$$

Observe that all constant polynomials $a \neq 0$ have degree 0. The 'zero' function is customarily assigned the degree $-\infty$. One reason is that with this convention, we have the degree function will have the property $deg\,(pq) = deg\,p + deg\,q$. [But in some other contexts, it may be convenient to assign some other value for $deg\,0$. For example if you take the set of all homogeneous polynomials of degree n in k-variables, then it would form a vector space, provided you assign the degree n to the zero polynomial.] All degree 1 polynomials are also referred to as *linear polynomials.* The **Fundamental Theorem of Algebra** (FTA) asserts that every *non-constant* polynomial assumes the value zero, i.e., the equation

$$p(z) = 0$$

has a solution. This is the same as saying that *every non constant polynomial has a root.* For an elementary proof of this theorem, see section 1.8. Later, in chapter 4, we shall see a proof of this theorem using Complex Analysis.

Now observe that if z_1 is a solution of $p(z) = 0$, then as in school algebra, we can perform division by $z - z_1$ and the remainder will be zero, i.e.,

$$p(z) = (z - z_1)q(z).$$

It also follows easily that $deg\ q(z) = n - 1$. Thus by repeated application of FTA, we can factorize $p(z)$ completely into linear factors and a constant:

$$p(z) = a_n(z - z_1)(z - z_2)\cdots(z - z_n), \quad a_n \neq 0. \tag{2.7}$$

We conclude that every polynomial of degree n has n roots. Also if $w \neq z_j, \quad j = 1, \ldots, n$, it is clear from (2.7) that $p(w) \neq 0$. Hence, $p(z)$ has precisely n roots. Observe that two or more of the roots z_j may coincide. If that is the case, we say, that the corresponding root is a *multiple root* with its *order* or *multiplicity* being equal to the number of times $z - z_j$ is repeated in the factorization (2.7). The factorization is unique up to a permutation of the factors.

Next, we observe that if z_1 is a repeated root then $p'(z_1) = 0$. Indeed if the multiplicity of z_1 is m in $p(z)$ then

$$p(z) = (z - z_1)^m q(z), \qquad (q(z_1) \neq 0)$$

which implies that,

$$p'(z) = m(z - z_1)^{m-1}q(z) + (z - z_1)^m q'(z) = (z - z_1)^{m-1}r(z),$$

where $r(z) = mq(z) + (z - z_1)q'(z)$. Since $r(z_1) \neq 0$, it follows that the multiplicity of z_1 in $p'(z)$ is $m - 1$.

As an entertaining exercise, let us prove the following theorem due to Gauss which has a lot of geometric content in it. Recall that if S is a subset of $\mathbb{C}$, then by *convex hull* of S we mean the set of all elements $\sum t_j s_j$, $s_j \in S$, where the sum is finite, $0 \leq t_j \leq 1$ and $\sum t_j = 1$.

Theorem 2.2.1 Gauss[1]*: Let $p(z)$ be a polynomial with complex coefficients. Then all roots of $p'(z)$ lie in the convex hull spanned by the roots of $p(z)$.*

[1]An equivalent version of this has been attributed to Lucas by Ahlfors.

Proof: For any complex number z, since $z\bar{z} = |z|^2$, it follows that for any $z \neq 0$, $\bar{z}^{-1}$ has the same argument as z. Let $z_1, \ldots, z_n$ be the roots of $p(z)$, so that $p(z) = c\prod_{j=1}^{n}(z - z_j)$. Then we have,

$$\frac{p'(z)}{p(z)} = \sum_{j=1}^{n} \frac{1}{z - z_j}.$$

Now suppose that w is a root of $p'(z)$. If $w = z_j$ for some j, there is nothing to prove. So, let $w \neq z_j$ for any j. Then it follows that

$$\sum_{j=1}^{n} \frac{1}{w - z_j} = 0 = \sum_{j=1}^{n} \frac{1}{\overline{\overline{w - z_j}}}. \tag{2.8}$$

On the other hand suppose w did not belong to the convex hull of $\{z_1, z_2, \ldots, z_n\}$, then it is easily seen that there is a straight line L passing through w such that all the points z_j lie strictly to one side of L. (Write full details as an exercise.) If L_1 is the line through the origin parallel to L, then it will mean that all the numbers $w - z_j$ lie on one side of L_1. Since $(\overline{w - z_j})^{-1}$ have the same argument as $w - z_j$, it follows that $(\overline{w - z_j})^{-1}$ also lie on the same side of L_1. But then, their sum cannot be zero! This contradiction to (2.8) proves the theorem. ♠

Remark 2.2.1 One can think of n forces of magnitude $|w - z_j|^{-1}$ acting on the point w and directed towards the point z_j. Then (2.8) can be interpreted as saying that the point w is at equilibrium under these forces. From this interpretation, the conclusion of the theorem is immediate for a physicist. That is how Gauss may have discovered this result. We have deliberately left out a few details in the above proof. These details should be supplied by the reader. (See exercise 23 in Misc. Exercises 1.9.)

(b) The rational functions: In (a), the domain of definition of our functions was the entire plane. We shall now define some $\mathbb{C}$-differentiable functions with their domains of definition not necessarily being the entire plane. These functions are of the form

$$\phi(z) = \frac{p(z)}{q(z)} \tag{2.9}$$

where p and q are polynomials, called *rational functions*. By canceling out common factors from both p and q, we can assume that p and q do not have any common factors. Then, obviously, $\phi(z)$ makes sense precisely when $q(z) \neq 0$ and so the domain of definition of $\phi(z)$ is $\mathbb{C} \setminus \{z \ : \ q(z) = 0\}$. We have by (2.4),

$$\phi'(z) = \frac{p'(z)q(z) - q'(z)p(z)}{(q(z))^2} \tag{2.10}$$

and so ϕ is $\mathbb{C}$-differentiable in its domain. Its derivative is another rational function having the same domain of definition. *Prove this statement. Caution: (2.10) may not be in the reduced form even though (2.9) is.* Later on, we shall have more opportunities to study these functions, particularly, the so called *fractional linear transformations,* which are of the form $\dfrac{az+b}{cz+d}$. At this stage, it may be worthwhile to note that the set of all polynomials in one variable with complex coefficients forms a commutative ring which we denote by $\mathbb{C}[z]$. One of the important property of this ring is that it is an *integral domain* , (i.e., a commutative ring in which product of two non zero elements is never zero). This follows easily from the 'unique factorization' property (2.7) that we have seen. The set of all rational functions forms a field, $\mathbb{C}(z)$, called the *field of fractions* of the integral domain $\mathbb{C}[z]$.

In order to get any other class of examples of complex differentiable functions, we have to use the power series. So, we prepare ourselves for this by recalling some basic facts about series in the next section.

Exercise 2.2

1. Show that composite of two polynomial (resp. rational) maps is a polynomial (resp. rational) map.

2. * Let $X_1, X_2, \ldots, X_n$ be n indeterminates. For each positive integer $k \leq n$, consider the polynomials:

$$\sigma_k = \sigma_k(X_1, \ldots, X_n) := \sum_{i_1 < \cdots < i_k} X_{i_1} X_{i_2} \cdots X_{i_k}.$$

They are called symmetric functions— for if you interchange X_i and X_j for any i, j the functions do not change. In fact, σ_k are called *elementary symmetric polynomials.* Of course, there are other symmetric polynomials. Another class of important symmetric polynomials are the *power sums*:

$$\rho_k = \rho_k(X_1, \ldots, X_n) := \sum_i X_i^k$$

defined for each $k \geq 1$. Establish the following Newton's identities:

$$\begin{aligned} \rho_1 - \sigma_1 &= 0 \\ \rho_2 - \sigma_1\rho_1 + 2\sigma_2 &= 0 \\ \cdots\cdots\cdots & \\ \rho_k - \sigma_1\rho_{k-1} + \cdots + (-1)^{k-1}\sigma_{k-1}\rho_1 + (-1)^k k\sigma_k &= 0 \end{aligned} \tag{2.11}$$

[Hint: Introduce the elaborate notation:

$$\sigma_k(n) = \sigma_k(X_1, \ldots, X_n) \text{ and } \rho_k(n) = \rho_k(X_1, \ldots, X_n).$$

Then observe that $\sigma_k(n) = \sigma_k(n-1) + X_n\sigma_{k-1}(n-1)$ and $\rho_k(n) = \rho_k(n-1) + X_n^k$. Now, induct on n to establish the above identities.] Show that any symmetric polynomial can be expressed as a linear combination of product of elementary symmetric polynomials, $\sigma_1, \ldots, \sigma_n$ in a unique way.

2.3 Analytic Functions: Power Series

We shall recall some basic results about the power series. In fact, we assume that the reader has some familiarity with this topic and hence our treatment of this topic is somewhat sketchy, as in the case of sequences and series.

Let $\mathbb{K} := \mathbb{R}$, or $= \mathbb{C}$. (Indeed $\mathbb{K}$ could be any field so far as we do not discuss convergence.)

Definition 2.3.1 By a *formal power series in one variable* t over $\mathbb{K}$, we mean a sum of the form

$$\sum_{n=0}^{\infty} a_n t^n, \quad a_n \in \mathbb{K}.$$

Let $\mathbb{K}[[t]]$ denote the set of all formal power series in t over $\mathbb{K}$. Observe that when at most a finite number of a_n are non zero the above sum gives a polynomial. Thus, all polynomials in t are power series in t, i.e., $\mathbb{K}[t] \subset \mathbb{K}[[t]]$.

Just like polynomials, we can add two power series 'term-by-term' and we can also multiply them by scalars, viz.,

$$\sum_n a_n t^n + \sum_n b_n t^n := \sum_n (a_n + b_n) t^n; \quad \alpha(\sum_n a_n t^n) := \sum_n \alpha a_n t^n.$$

Verified that the above two operations make $\mathbb{K}[[t]]$ into a vector space over $\mathbb{K}$.

Further, we can even multiply two formal power series:

$$\left(\sum_n a_n t^n\right)\left(\sum_n b_n t^n\right) := \sum_n c_n t^n,$$

where, $c_n = \sum_{k=0}^{n} a_k b_{n-k}$. This product is called the *Cauchy product.*

One can directly check that $\mathbb{K}[[t]]$ is then a commutative ring with the multiplicative identity being the power series

$$1 := \sum_n a_n t^n$$

where, $a_0 = 1$ and $a_n = 0, n \geq 1$. Together with the vector space structure, $\mathbb{K}[[t]]$ is actually a $\mathbb{K}$-algebra.) Observe that the ring of polynomials in t forms a subring of $\mathbb{K}[[t]]$. What we are now interested in is to get nice functions out of power series.

Observe that, if $p(t)$ is a polynomial over $\mathbb{K}$ then by the **method of substitution,** it defines a function $a \mapsto p(a)$, from $\mathbb{K}$ to $\mathbb{K}$. It is customary to denote this map by $p(t)$ itself. However, due to the infinite nature of the sum involved, given a power series P and a point $a \in \mathbb{K}$, the substitution $P(a)$ may not make sense in general. This is the reason why we have to treat power series with a little more care, via the notion of convergence.

Definition 2.3.2 A formal power series $P(t) = \sum_n a_n z^n$ is said to be *convergent at* $z_0 \in \mathbb{C}$ if the sequence $\{s_n\}$, where, $s_n = \displaystyle\sum_{k=0}^{n} a_k z_0^k$ is convergent. In that case we write

$$P(z_0) = \lim_{n\to\infty} s_n$$

for this limit. Putting $t_n = a_n z_0^n$, this just means that the series of complex numbers $\sum_n t_n$ is convergent.

Remark 2.3.1 Observe that every power series is convergent at 0.

Definition 2.3.3 A power series is said to be a convergent power series, if it is convergent at some point $z_0 \neq 0$.

The following few theorems, which are attributed to Cauchy-Hadamard[2] and Abel[3], are most fundamental in the theory of convergent power series.

[2]Jacques Hadamard(1865-1963) was a French Mathematician who was the most influential mathematician of his days, worked in several areas of mathematics such as complex analysis, analytic number theory, partial differential equations, hydrodynamics and logic.

[3]Niels Henrik Abel (1802-1829) was a Norwegian, who died young under deprivation. At the age of 21, he proved the impossibility of solving a general quintic by radicals. He did not get any recognition during his life time for his now famous works on convergence, on so called abelian integrals, and on elliptic functions.

Theorem 2.3.1 Cauchy-Hadamard Formula: *Let* $P = \sum_{n\geq 0} a_n t^n$ *be a power series over* $\mathbb{C}$. *Put* $L = \limsup_n \sqrt[n]{|a_n|}$ *and* $R = \frac{1}{L}$ *with the convention* $\frac{1}{0} = \infty$; $\frac{1}{\infty} = 0$. *Then*
(a) *for all* $0 < r < R$, *the series* $P(t)$ *is absolutely and uniformly convergent in* $|z| \leq r$ *and*
(b) for all $|z| > R$, *the series is divergent.*

Proof: (a) Let $0 < r < R$. Choose $r < s < R$. Then $1/s > 1/R = L$ and hence by property (Limsup-I) (see definition 1.4.2), we must have n_0 such that for all $n \geq n_0$, $\sqrt[n]{|a_n|} < 1/s$. Therefore, for all $|z| \leq r$, $|a_n z^n| < (r/s)^n, \quad n \geq n_0$. Since $r/s < 1$, by Weierstrass majorant criterion, (Theorem 1.4.8), it follows that $P(z)$ is absolutely and uniformly convergent.
(b) Suppose $|z| > R$. We fix s such that $|z| > s > R$. Then $1/s < 1/R = L$, and hence by property (Limsup-II), there exist infinitely many n_j, for which $\sqrt[n_j]{|a_{n_j}|} > 1/s$. This means that $|a_{n_j} z^{n_j}| > (|z|/s)^{n_j} > 1$. It follows that the n^{th} term of the series $\sum_n a_n z^n$ does not converge to 0 and hence the series is divergent. ♠

Definition 2.3.4 Given a power series $\sum_n a_n t^n$,

$$R = sup\{|z| : \sum_n a_n z^n < \infty\}$$

is called the *radius of convergence* of the series. The above theorem gives you the formula for R.

Remark 2.3.2 Observe that if $P(t)$ is convergent for some z, then the radius of convergence of P is at least $|z|$. The second part of the theorem gives you the formula for it. This is called the **Cauchy-Hadamard formula.** It is implicit in this theorem that the the collection of all points at which a given power series converges consists of an open disc centered at the origin and perhaps some points on the boundary of the disc. This disc is called the disc of convergence of the power series. Observe that the theorem does not say anything about the convergence of the series at points on the boundary $|z| = R$. The examples below will tell you that any thing can happen on the boundary.

Example 2.3.1 The series $\displaystyle\sum_n t^n, \sum_n \frac{t^n}{n}, \sum_n \frac{t^n}{n^2}$ all have radius of convergence 1. The first one is not convergent at any point of the boundary of the disc of convergence $|z| = 1$. The second is convergent at all the points of the boundary except at $z = 1$ (Dirichlet's test) and the last one is convergent at all the points of the boundary (compare with $\sum_n \frac{1}{n^2}$). These examples clearly illustrate that the boundary behavior of a power series needs to be studied more carefully.

Remark 2.3.3 It is not hard to see that the sum of two convergent power series is convergent. Indeed, the radius of convergence of the sum is at least the minimum of the radii of convergence of the summands. Similar statement holds for Cauchy product. Since Cauchy product of two convergent series with non negative real coefficients is convergent, it follows that the radius of convergence of the Cauchy product of two series is at least the minimum of the radii of convergence of the two series.

Example 2.3.2 Here is an example of usefulness of Cauchy's product. Consider the geometric series $g(t) = 1 + t + t^2 + \cdots$ with radius of convergence equal to 1. We can easily compute $(g(t))^2$ and see that

$$(g(t))^2 = 1 + 2t + 3t^2 + \cdots + nt^{n-1} + \cdots$$

which also should have radius convergence at least 1. Also it is not convergence at 1. Hence the radius of convergence is exactly one. Thus, it follows that $\sum_k kt^k = tg(t)^2$ also has radius of convergence equal to 1. By Cauchy Hadamard's theorem, it follows that $\limsup_n \sqrt[n]{n} = 1$. In turn, it follows that for all integers m, the series $\sum_k k^m t^k$ have radius of convergence 1.

Definition 2.3.5 Given a power series $P(t) = \sum_{n\geq 0} a_n t^n$, the derived series $P'(t)$ is defined by taking term-by-term differentiation: $P'(t) = \sum_{n\geq 1} n a_n t^{n-1}$. The series $\sum_{n\geq 0} \frac{a_n}{n+1} t^{n+1}$ is called the integrated series.

As an application of Cauchy-Hadamard formula, we derive:

Theorem 2.3.2 *A power series $P(t)$, its derived series $P'(t)$ and any series obtained by integrating $P(t)$ all have the same radius of convergence.*

Proof: Let the radius of convergence of $P(t) = \sum_n a_n t^n$, and $P'(t)$ be r, r' respectively. It is enough to prove that $r = r'$.

We will first show that $r \geq r'$. For this we may assume without loss of generality that $r' > 0$. Let $0 < r_1 < r'$. Then

$$\sum_{n\geq 1} |a_n| r_1^n = r_1 \left(\sum_{n\geq 1} n |a_n| r_1^{n-1} \right) < \infty.$$

It follows that $r \geq r_1$. Since this is true for all $0 < r_1 < r'$ this means $r \geq r'$.

Now to show that $r \le r'$, we can assume that $r > 0$ and let $0 < r_1 < r$. Choose r_2 such that $r_1 < r_2 < r$. Then for each $n \ge 1$

$$nr_1^{n-1} \le \frac{n}{r_1}\left(\frac{r_1}{r_2}\right)^n r_2^n \le \frac{M}{r_1} r_2^n$$

where $M = \sum_{k\ge 1} k\left(\frac{r_1}{r_2}\right)^k < \infty$, since the radius of convergence of $\sum_k kt^k$ is at least 1 (See Example 2.3.2.) Therefore,

$$\sum_{n\ge 1} n|a_n|r_1^{n-1} \le \frac{M}{r_1}\sum_{n\ge 1}|a_n|r_2^n < \infty.$$

We conclude that $r' \ge r_1$ and since this holds for all $r_1 < r$, it follows that $r' \ge r$. ♠

Remark 2.3.4

(i) For any sequence $\{b_n\}$ of non negative real numbers, one can directly try to establish

$$\limsup_n \sqrt[n]{(n+1)b_{n+1}} = \limsup_n \sqrt[n]{b_n}$$

which is equivalent to proving theorem 2.3.2. However, the full details of such a proof are no simpler than the above proof. In any case, this way, we would not have got the limit of these derived series.

(ii) A power series with radius of convergence 0 is apparently 'useless for us', for it only defines a function at a point. It should noted that in other areas of mathematics, there are many interesting applications of formal power series which need be convergent,

(iii) A power series $P(t)$ with a positive radius of convergence R defines a continuous function $z \mapsto p(z)$ in the disc of convergence $B_R(0)$, by theorem 1.4.9. Also, by shifting the origin, we can even get continuous functions defined in $B_R(z_0)$, viz., by substituting $t = z - z_0$.

(iv) One expects that functions which agree with a convergent power series in a small neighborhood of every point will have properties akin to those of polynomials. So, the first step towards this is to see that a power series indeed defines a $\mathbb{C}$-differentiable function in the disc of convergence.

Theorem 2.3.3 Abel: *Let $\sum_{n\ge 0} a_n t^n$ be a power series of radius of convergence $R > 0$. Then the function defined by*

$$f(z) = \sum_n a_n (z - z_0)^n$$

is complex differentiable in $B_r(z_0)$. Moreover the derivative of f is given by the derived series

$$f'(z) = \sum_{n\geq 1} n a_n (z - z_0)^{n-1}$$

inside $|z - z_0| < R$.

Proof: Without loss of generality, we may assume that $z_0 = 0$. We already know that the derived series is convergent in $B_R(0)$ and hence defines a continuous function g on it. We have to show that this function g is the derivative of f at each point of $B_R(0)$. So, fix a point $z \in B_R(0)$. Let $|z| < r < R$ and let $0 \neq |h| \leq r - |z|$ so that $|z + h| \leq r$. Consider the difference quotient

$$\frac{f(z+h) - f(z)}{h} - g(z) = \sum_{n\geq 1} u_n(h) \tag{2.12}$$

where, we have put $u_n(h) := \dfrac{a_n[(z+h)^n - z^n]}{h} - n a_n z^{n-1}$. We must show that given $\epsilon > 0$, there exists $\delta > 0$ such that for all $0 < |h| < \delta$, we have,

$$\left| \frac{f(z+h) - f(z)}{h} - g(z) \right| < \epsilon. \tag{2.13}$$

The idea here is that the sum of first few terms can be controlled by continuity whereas the remainder term can be controlled by the convergence of the derived series. Using the algebraic formula

$$\frac{\alpha^n - \beta^n}{\alpha - \beta} = \sum_{k=0}^{n-1} \alpha^{n-1-k}\beta^k,$$

putting $\alpha = z + h, \beta = z$ we get,

$$u_n(h) = a_n[(z+h)^{n-1} + (z+h)^{n-2}z + \cdots + (z+h)z^{n-2} + z^{n-1} - nz^{n-1}]. \tag{2.14}$$

Since $|z| < r$ and $|z + h| < r$, it follows that

$$|u_n(h)| \leq 2n|a_n|r^{n-1}. \tag{2.15}$$

Since the derived series has radius of convergence $R > r$, it follows that we can find n_0 such that

$$\left| \sum_{n\geq n_0} u_n(h) \right| \leq 2 \sum_{n\geq n_0} |a_n| n r^{n-1} < \epsilon/2. \tag{2.16}$$

On the other hand, again using (2.14), each $u_n(h)$ is a polynomial in h which vanishes at $h = 0$. Therefore so does the finite sum $\sum_{n<n_0} u_n(h)$. Hence by continuity, there exists $\delta' > 0$ such that for $|h| < \delta'$ we have,

$$\left|\sum_{n<n_0} u_n(h)\right| < \epsilon/2. \tag{2.17}$$

Taking $\delta = min\{\delta', r - |z|\}$ and combining (2.16) and (2.17) yields (2.13). ♠

Definition 2.3.6 Let $f : \Omega \longrightarrow \mathbb{C}$ be a function. We say f is *analytic* at $z_0 \in \Omega$ if there exists $r > 0$ and a power series P of radius of convergence $\geq r$ such that $B_r(z_0) \subset \Omega$ and $f(z) = P(z - z_0)$ for all $z \in B_r(z_0)$. We say f is analytic in Ω if it is so at each point $z_0 \in \Omega$.

As seen above it follows that an analytic function is $\mathbb{C}$-differentiable any number of times. On the other hand, what is not obvious is that if $P(t)$ is a formal power series with positive radius of convergence r, then the function $f(z) = P(z - z_0)$ is analytic in $|z - z_0| < r$. (Observe that the definition only says that it is analytic *at* z_0.) This fact can be directly proved (See the proposition below). [However, you may skip this, as the result will follow easily from Taylor's theorem which we shall prove in chapter 4.]

Proposition 2.3.1 Let $P(t) = \sum_{n=0}^{\infty} a_n t^n$ be a power series with radius of convergence $\rho > 0$. Let z_0 be any point such that $|z_0| < \rho$. Let $P^{(n)}(z_0)$ denote the n^{th} derivative of $P(z)$ at z_0. Then the power series

$$Q(t) = \sum_{n=0}^{\infty} \frac{P^{(n)}(z_0)}{n!} t^n$$

has radius of convergence at least $\rho - |z_0|$. Moreover, for all $|w - z_0| < \rho - |z_0|$, we have

$$P(w) = Q(w - z_0). \tag{2.18}$$

In particular, the function $z \mapsto P(z)$ is analytic at every point of the disc $|z| < \rho$.

Proof: Note that by repeated application of theorem 2.3.3, $P^{(n)}(z_0)$ makes sense. Consider the double series

$$\sum_{m\geq 0, n\geq 0} \frac{(m+n)!}{m!n!} a_{m+n} z_0^n (w - z_0)^m. \tag{2.19}$$

By formally summing it up in two different ways we obtain both LHS and RHS of (2.18). Therefore, we need only to justify the rearrangements. This will follow if we show that the double series (2.19) is absolutely convergent for $|w - z_0| < \rho - |z_0|$.

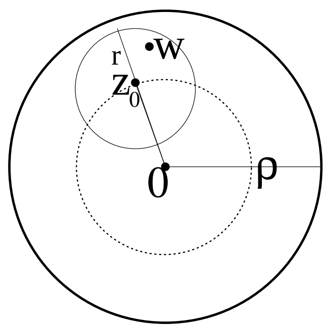


Fig.11

Choose r such that $|w - z_0| < r < \rho - |z_0|$. Then $r < \rho$ and hence

$$\begin{aligned}\sum_{m\geq 0}\sum_{n\geq 0}\frac{(m+n)!}{m!n!}|a_{m+n}||z_0|^n|w-z_0|^m &\leq \sum_{m\geq 0}\sum_{n\geq 0}\frac{(m+n)!}{m!n!}|a_{m+n}||z_0|^n|r^m \\ &= \sum_{l\geq 0}|a_l|(r+|z_0|)^l < \infty.\end{aligned}$$

♠

Remark 2.3.5

(i) The above theorem, also known as Taylor's theorem has nothing to do with complex numbers in the sense that it is true for power series with real coeffients and even the proof is the same. In that case, we obtain what are known as real analytic functions. They share several properties common with complex analytic functions.

(ii) It follows that an analytic function is repeatedly differentiable, with its n^{th} derivative at z_0 given by $n!a_n$. As a consequence, we know that, at each point the power series representing the function is unique.

(iii) It is fairly obvious that the sum of any two analytic functions is again an analytic function. The corresponding statement about power series is that the sum of two formal power series is convergent with radius of convergence at least the minimum of the two radii of convergence. Similarly, the product of two analytic functions is also analytic. The identity function written $f(z) = z$ is clearly analytic in the entire plane (take $P(t) = z_0 + t$ to see that f is analytic at z_0). Starting from this and using the above two observations we can deduce that any polynomial function is analytic throughout the plane.

(iv) Later on we shall prove that every complex differentiable function in an open set is analytic, thus completing the picture. In particular, this will then prove that a function which is complex differentiable once is complex differentiable infinitely many times. Of course, such a result is far from being true in the real differentiable case. Examples such as $g(x) = x^n|x|$ illustrate the existence of functions which are n-times real differentiable but not $(n+1)$-times. Also, there are functions of a real variable which are $\mathcal{C}^\infty$ (i.e., differentiable any number of times) but not real analytic. (Take $f(x) = 0$ for $x \le 0$ and $= e^{-1/x^2}$ for $x > 0$.)

Remark 2.3.6 We have remarked earlier that a power series with radius of convergence 0 is apparently useless for us. This may be so, so far as we are interested in getting analytic functions out of them. Even though we have introduced power series for this specific purpose, we cannot ignore completely their usefulness elsewhere. Indeed in combinatorics, many useful power series are all of radius of convergence 0. The radius of convergence of power series is the last thing one would be bothered about there. The most important property of power series is that it allows 'unrestricted' algebraic manipulation. We shall be satisfied with a refreshingicing and illustrative example below and will not pursue this line of study any further. For the time being, we are more interested in getting plenty of examples of $\mathbb{C}$-differentiable functions via analytic functions. That is the topic that we are going to take up in the next section.

Example 2.3.3 Hemachandra Numbers

When you are creating some rhythmic patterns on a drum, you make essentially a sequence of sounds of one or two syllables such as in 'DHAA(1)- KI(1)-TA(1)'. Here DHAA is called a long beat counting for two units of time whereas KI and TA are short beats each counting for a single unit of time. Thus the total time duration in the above rhythm is 4.

For any positive integer n, let the n^{th} **Hemechandra number** H_n denote the number of possible rhythms of total time duration n that one can make. Clearly $H_1 = 1$ and $H_2 = 2$. To determine H_n Hemachandra [4] simply remarks that the last syllable is either of one beat or two beats. This is clearly his way of giving the formula

$$H_n = H_{n-1} + H_{n-2} \quad \text{for all } n \ge 3.$$

[4]Hemachandra Suri (1089-1175) was born in Dhandhuka, Gujarat. He was a Jain monk and was an adviser to king Kumarapala. His work in early 11 century is already based on even earlier works of Gopala.

Obviously, these numbers were known to Indian poets, musicians and practicing percussionists long before Hemachandra.

Define $F_0 = 0, F_1 = 1$ and $F_n = F_{n-1} + F_{n-2}, n \geq 2$. Note that $F_n = H_{n-1}, n \geq 2$. These F_n are called **Fibonacci numbers.** [5] (Thus the first few Fibonacci numbers are $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$.)

Form the formal power series

$$F(z) = \sum_{n=0}^{\infty} F_n z^n \tag{2.20}$$

Multiplying the given recurrence relation by t^n and summing over from 2 to ∞ gives

$$\sum_{n=2}^{\infty} F_n t^n = t \sum_{n=2}^{\infty} F_{n-1} t^{n-1} + t^2 \sum_{n=2}^{\infty} F_{n-2} t^{n-2} \tag{2.21}$$

and hence

$$(1 - t - t^2)F(t) = t.$$

Write $(1 - t - t^2) = (1 - \alpha t)(1 - \beta t)$, where $\alpha = \dfrac{1+\sqrt{5}}{2}$ and $\beta = \dfrac{1-\sqrt{5}}{2}$. Put $S_w(t) = 1 + wt + w^2 t^2 + \cdots$. Then $(1 - wt)S_w(t) = 1$ and

$$F(t) = S_\alpha(t) S_\beta(t).$$

Comparing the coefficients of t^n on either side, we get,

$$F_{n+1} = \sum_{j=0}^{n} \alpha^j \beta^{n-j} = \frac{\alpha^{n+1} - \beta^{n+1}}{\alpha - \beta} = \frac{1}{\sqrt{5}}(\alpha^{n+1} - \beta^{n+1}) \tag{2.22}$$

Exercise 2.3 *

1. Verify that $\mathbb{K}[[t]]$ is a $\mathbb{K}$-algebra, i.e., a $\mathbb{K}$-vector space which is a commutative ring with a multiplicative unit.

2. For a non zero element $p = \sum_{n\geq 0} a_n t^n \in \mathbb{K}[[t]]$, the order $\omega(p)$ of p is defined to be the least integer for which $a_n \neq 0$. By convention, we define $\omega(0) = +\infty$. (This is consistent with the convention that infimum of an empty subset of real numbers is $+\infty$.) Show that $\omega(p + q) \geq min\{\omega(p), \omega(q)\}$ and $\omega(pq) = \omega(p) + \omega(q)$.

[5] Leonardo Pisano (Fibonacci) was born in Pisa, Italy (1175-1250) whose book Liber abbaci introduced the Hindu-Arabic decimal system to the western world. He discovered these numbers at least 50 years later than Hemachandra's record.

3. Given $p \in \mathbb{K}[[t]]$, show that p has a multiplicative inverse iff $\omega(p) = 0$.

4. Show that $\mathbb{K}[[t]]$ is an integral domain, i.e., $p, q \in \mathbb{K}[[t]]$ such that $pq = 0$ implies $p = 0$ or $q = 0$.

5. A family $\{p_j = \sum_n a_{n,j} t^n\}$ of elements in $\mathbb{K}[[t]]$ is said to be a *summable family* if for each $n \geq 0$ the number of j for which the coefficient of t^n in p_j is not zero is finite, i.e.,
$$\#\{j \ : \ a_{n,j} \neq 0\} < \infty.$$
In this case, we define the sum of this family to be the element $p(t) = \sum_{n \geq 0} a_n t^n$ where $a_n = \sum_j a_{n,j}$.

   Put $p = \sum_n a_n t^n, q = \sum_n b_n t^n$.

   (a) Verify that the Cauchy product pq is indeed the sum of the family $\{a_n b_m t^{m+n}\}$.

   (b) If $\{p_j\}$ is a summable series then for any series q the family $\{p_j q\}$ is also summable.

   (c) Assume that $b = 0$, i.e., $\omega(q) \geq 1$. Then show that the family $\{a_n q^n \ : \ n \geq 0\}$ is summable.

6. The sum of the above family of series in (c) is called the series obtained by **substituting $t = q$ in p** or **the composition series** and written $p \circ q$. Continue to assume that $b_0 = 0$. Let $p \circ q(t) = \sum_n \alpha_n t^n$.

   (a) Show that for each positive integer n, there exist a (universal) polynomial $U_n(A_1, \ldots, A_n, B_1, B_2, \ldots, B_n)$ with the following properties:
   (i) all coefficients are positive integers;
   (ii) Each U_n is linear in $A_0, A_1, \ldots, A_n$, and B_n with coefficient of $B_n = A_1$.
   (iii) Each U_n is weighted homogeneous of degree $n + 1$ where
   $$\deg A_j = 1; \quad \deg B_j = j.$$
   Moreover, U_n have the property
   $$\alpha_n = U_n(a_1, \ldots, a_n, b_1, b_2, \ldots, b_n). \tag{2.23}$$
   Write down explicitly U_1, U_2, U_3.

   (b) Show that $(p_1 + p_2) \circ q = p_1 \circ q + p_2 \circ q$.

(c) $(p_1p_2)\circ q = (p_1\circ q)(p_2\circ q)$.

(d) If $r = \sum_n c_nt^n$ is such that $c_0 = 0$, then we have

$$p\circ(q\circ r) = (p\circ q)\circ r.$$

(e) Consider the element $I(t) = t \in \mathbb{K}[[t]]$. Show that it is a two-sided identity for the composition, i.e., $p\circ I = I\circ p = p$ for all $p \in \mathbb{K}[[t]]$.

7. Show that if p is a polynomial then for any $q \in \mathbb{K}[[t]]$, the composition $p\circ q$ makes sense for all $q \in \mathbb{K}[[t]]$, i.e., even without the assumption that $\omega(q) \geq 1$.

8. Let $'$ denote the derived series. Show that

   (a) $p' = 0$ iff p is a constant;

   (b) $(p+q)' = p' + q'$; $(pq)' = p'q + pq'$.

   (c) $(p^n)' = np^{n-1}p'$, for all integers n. (Here if n is negative you have to assume that p^n makes sense, which is guaranteed if $p(0) \neq 0$.)

   (d) If $\{p_j\}$ is a summable family of power series then $(\sum_j p_j)' = \sum_j p_j'$.

   (e) **Chain rule** $(p\circ q)' = (p'\circ q)q'$.

9. **Inverse Function Theorem for Formal Power Series** Given an element $p = \sum_{n\geq 0} a_nt^n \in \mathbb{K}[[t]]$, show that there is a $q \in \mathbb{K}[[t]]$ such that $q(0) = 0$ and $p\circ q = I$ iff $a_0 = 0$ and $a_1 \neq 0$. Show that such a q is unique. Further, in this case, show that and $q\circ p = I$.

10. In the above exercise, if $a_0 \neq 0$, we can still do something, viz., we consider $r = p - a_0$, apply the above conclusion to r to get s such that $s\circ r = I = r\circ s$; $r(0) = 0$. From this we conclude that $p\circ s(t) = r\circ s(t) + a_0 = t + a_0$;

11. Let us consider two of the most important series

$$E(t) = 1 + t + \frac{t^2}{2!}\cdots + \frac{t^n}{n!} + \cdots$$

$$L(t) = t - \frac{t^2}{2} + \frac{t^3}{3} + - \cdots + (-1)^{n-1}\frac{t^n}{n} + - \cdots$$

respectively called the **exponential series** and **logarithmic series.**

   (a) Verify that $E(t+s) = E(t)E(s)$;

(b) $E'(t) = E(t)$;

(c) By Ex. 10 there is a unique $F \in \mathbb{K}[[t]]$ such that $F(0) = 0$ and $L \circ F = Id = F \circ L$. Verify that $F = E - 1$. Conclude that $E \circ L(t) = 1 + t$ and $L \circ (E(t) - 1) = t$. We put $Ln(1+t) = L(t)$ so that we have $E \circ Ln = Id$ and $Ln \circ E = L \circ (E(t) - 1) = t$.

Exercises on Convergent Power Series

Throughout these exercises let $p(t) = \sum_{n=0}^{\infty} a_n t^n, q(t) = \sum_{n=0}^{\infty} b_n t^n$ be two power series.

12. Let p, q both have radius of convergence $\geq r > 0$.

    (a) The radii of convergence of both $p + q$ and pq are $\geq r$. Moreover for $|z| < r$, we have $(p + q)(z) = p(z) + q(z)$; and $(pq)(z) = p(z)q(z)$.

    (b) Assume further that $q(0) = b_0 = 0$. Then the composite series $p \circ q$ has positive radius of convergence.

13. Let $pq = 1$. If the radius of convergence of p is positive then so is the radius of convergence of q.

14. Given $\alpha \neq 0, \beta \neq 0$, and a positive integer n, show that there is a unique formal power series p such that $p(0) = \alpha$, and $p^n = \alpha^n + \beta t$. Show that p is of positive radius of convergence.

15. Given $\alpha \neq 0$, show that there is a unique power series p of positive radius of convergence such that

$$p^2 = \alpha^2 + \beta t + \gamma t^2; \quad p(0) = \alpha.$$

16. Show that there is a unique power series which satisfies

$$p^2 - (\alpha^2 + \beta t)p + \gamma t = 0; \quad p(0) = 0 \tag{2.24}$$

and it has a positive radius of convergence.

17. For some positive numbers α, r, M, let

$$P(t) = \alpha t - \sum_{n \geq 2} \frac{M}{r^n} t^n.$$

If Q is the compositional inverse of P, show that Q is of positive radius of convergence.

18. Let $p(t) = \sum_n a_n t^n, q(t) = \sum_n b_n t^n, a_0 = 0 = b_0, a_1 \neq 0$ and $p \circ q = Id$. Suppose $P(t) = A_1 t - \sum_{n \geq 2} A_n t^n$ is such that $A_1 = |a_1|$, and $|a_n| < A_n$ for all $n \geq 2$. Let $Q = \sum_n B_n T^n$ be the compositional inverse of P with $Q(0) = 0$. Then show that $|b_n| \leq B_n$ for all n.

19. **Inverse Function Theorem for Analytic Functions** Let $p \circ q = Id$, where $p(0) = 0$ and $p'(0) \neq 0$. If p is of positive radius of convergence then so is q.

2.4 The Exponential and Trigonometric Functions

The exponential function plays a central role in analysis, more so in the case of complex analysis and is going to be our first example using the power series method. We define

$$\boxed{\exp z := e^z := \sum_{n \geq 0} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .} \tag{2.25}$$

By comparison test it follows that for any real number $r > 0$, the series $\exp(r)$ is convergent. Therefore, the radius of convergence of (2.25) is ∞. Hence from theorem 2.3.3, we have, exp is differentiable throughout $\mathbb{C}$ and its derivative is given by

$$\boxed{\exp'(z) = \sum_{n \geq 1} \frac{n}{n!} z^{n-1} = \exp(z)} \tag{2.26}$$

for all z. It may be worth recalling some elementary facts about the exponential function that you probably know already. Let us denote by

$$e := \exp(1) = 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots$$

Clearly, $\exp(0) = 1$ and $2 < e$. By comparing with the geometric series $\sum_n \frac{1}{2^n}$, it also follows that $e < 3$. An interesting expression for e is:

$$\boxed{e = \lim_{n \longrightarrow \infty} \left(1 + \frac{1}{n}\right)^n .} \tag{2.27}$$

To see this, put $t_n = \sum_{k=0}^n \frac{1}{k!}, \quad s_n = \left(1 + \frac{1}{n}\right)^n$, use binomial expansion to see that

$$\limsup_n s_n \leq e \leq \liminf_n s_n.$$

Moreover, since $\overline{\sum_0^n \frac{z^k}{k!}} = \sum_0^n \frac{\bar{z}^k}{k!}$, by continuity of the conjugation, it follows that

$$\boxed{\boxed{\overline{\exp z} = \exp \bar{z}.}} \tag{2.28}$$

Next, formula (2.26) together with the property $\exp(0) = 1$, tells us that $\exp$ is a solution of the initial value problem:

$$\boxed{\boxed{f'(z) = f(z); \quad f(0) = 1.}} \tag{2.29}$$

It can be easily seen that any analytic function which is a solution of (2.29) has to be equal to $\exp$. (Ex. Prove this directly without quoting uniqueness of solution of initial value problem.)

We can verify that

$$\boxed{\boxed{\exp(a+b) = \exp(a)\exp(b), \quad \forall a, b \in \mathbb{C}}} \tag{2.30}$$

directly by using the product formula for power series. (Use binomial expansion of $(a+b)^n$; Ex. 2.3.11(a).) This can also be proved by using the uniqueness of the solution of (2.29) which we shall leave it you as an entertaining exercise (2.5.13).

Thus, we have shown that $\exp$ defines a homomorphism from the additive group $\mathbb{C}$ to the multiplicative group $\mathbb{C}^\star := \mathbb{C} \setminus \{0\}$. As a simple consequence of this rule we have, $\exp(nz) = \exp(z)^n$ for all integers n. In particular, we have, $\exp(n) = e^n$. This is the justification to have the notation

$$\boxed{\boxed{e^z := \exp(z).}}$$

Combining (2.28) and (2.30), we obtain,

$$|e^{\imath y}|^2 = e^{\imath y}\overline{e^{\imath y}} = e^{\imath y}e^{-\imath y} = e^0 = 1.$$

Hence,

$$\boxed{\boxed{|e^{\imath y}| = 1, \quad y \in \mathbb{R}.}} \tag{2.31}$$

Example 2.4.1 Trigonometric Functions. Recall the two Taylor series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \cdots; \qquad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \cdots,$$

valid on the entire of $\mathbb{R}$, since the radii of convergence of the two series are ∞. Motivated by this, we can define the complex trigonometric functions by

$$\boxed{\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - + \cdots ; \quad \cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - + \cdots .} \tag{2.32}$$

Check that

$$\boxed{\sin z = \frac{e^{\imath z} - e^{-\imath z}}{2\imath}; \quad \cos z = \frac{e^{\imath z} + e^{-\imath z}}{2}.} \tag{2.33}$$

It turns out that these complex trigonometric functions also have differentiability properties similar to the real case, viz., $(\sin z)' = \cos z; (\cos z)' = -\sin z$, etc.. Also, from (2.33) additive properties of sin and cos can be derived. For instance, an immediate consequence of (2.33) is that:

$$\boxed{\sin^2 z + \cos^2 z = 1}$$

valid for all $z \in \mathbb{C}$.

Other trigonometric functions are defined in terms of sin and cos as usual. For example, we have $\tan z = \dfrac{\sin z}{\cos z}$ and its domain of definition is all points in $\mathbb{C}$ at which $\cos z \neq 0$.

In what follows, we shall obtain other properties of the exponential function by the formula

$$\boxed{e^{\imath z} = \cos z + \imath \sin z.} \tag{2.34}$$

In particular,

$$\boxed{e^{x+\imath y} = e^x e^{\imath y} = e^x(\cos y + \imath \sin y).} \tag{2.35}$$

It follows that $e^{2\pi\imath} = 1$. Indeed, we shall prove that $e^z = 1$ iff $z = 2n\pi\imath$, for some integer n. Observe that $e^x \geq 0$ for all $x \in \mathbb{R}$ and that if $x > 0$ then $e^x > 1$. Hence for all $x < 0$, we have, $e^x = 1/e^{-x} < 1$. It follows that $e^x = 1$ iff $x = 0$. Let now $z = x + \imath y$ and $e^z = 1$. This means that $e^x \cos y = 1$ and $e^x \sin y = 0$. Since $e^x \neq 0$ for any x, we must have, $\sin y = 0$. Hence, $y = m\pi$, for some integer m. Therefore $e^x \cos m\pi = 1$. Since $\cos m\pi = \pm 1$ and $e^x > 0$ for all $x \in \mathbb{R}$, it follows that $\cos m\pi = 1$ and $e^x = 1$. Therefore $x = 0$ and $m = 2n$, as desired.

Finally, let us prove:

$$\boxed{\exp(\mathbb{C}) = \mathbb{C}^\star.} \tag{2.36}$$

Write $0 \neq w = r(\cos\theta + \imath \sin\theta), r \neq 0$. Since e^x is a monotonically increasing function and has the property $e^x \longrightarrow 0$, as $x \longrightarrow -\infty$ and $e^x \longrightarrow \infty$ as $x \longrightarrow \infty$, it follows from Intermediate Value Theorem that there exist x such that $e^x = r$. (Here x is nothing but $\ln r$.) Now take $y = \theta, z = x + \imath\theta$ and use (2.35) to verify that $e^z = w$. This is one place, where we are heavily depending on the intuitive properties of the angle and the corresponding properties of the real sin and cos functions. We remark that it is possible to avoid this by defining sin and cos by the formula (2.33) in terms of exp and derive all these properties rigorously from the properties of exp alone. (See Exercise 2.5.15)

Remark 2.4.1 One of the most beautiful equations:

$$\boxed{\boxed{e^{\pi\imath} + 1 = 0}} \tag{2.37}$$

which relates in a simple arithmetic way, *five* of the most fundamental numbers, made Euler[6] believe in the existence of God!

Example 2.4.2 Let us study the mapping properties of tan function. Since $\tan z = \frac{\sin z}{\cos z}$, it follows that tan is defined and complex differentiable at all points where $\cos z \neq 0$. Also, $\tan(z + n\pi\imath) = \tan z$. In order to determine the range of this function, we have to take an arbitrary $w \in \mathbb{C}$ and try to solve the equation $\tan z = w$ for z. Putting $e^{\imath z} = X$, temporarily, this equation reduces to $\dfrac{X^2 - 1}{\imath(X^2 + 1)} = w$. Hence $X^2 = \dfrac{1 + \imath w}{1 - \imath w}$. This latter equation makes sense, iff $w \neq -\imath$ and then it has, in general two solutions. The solutions are $\neq 0$ iff $w \neq \imath$. Once we pick such a non zero X we can then use the ontoness of $\exp : \mathbb{C} \longrightarrow \mathbb{C} \setminus \{0\}$, to get a z such that that $e^{\imath z} = \pm X$. It then follows that $\tan z = w$ as required. Therefore we have proved that the range of tan is equal to $\mathbb{C} \setminus \{\pm\imath\}$. From this analysis, it also follows that $\tan z_1 = \tan z_2$ iff $z_1 = z_2 + n\pi\imath$.

Likewise, the hyperbolic functions are defined by

$$\boxed{\boxed{\sinh z = \frac{e^z - e^{-z}}{2}; \quad \cosh z = \frac{e^z + e^{-z}}{2}.}} \tag{2.38}$$

It is easy to see that these functions are $\mathbb{C}$-differentiable. Moreover, all the usual identities which hold in the real case amongst these functions also hold in the complex case and can be verified directly. One can study the mapping properties of these functions as well, which have wide range of applications.

[6]See E.T. Bell's book for some juicy stories.

Remark 2.4.2 Before we proceed onto another example, we would like to draw your attention to some special properties of the exponential and trigonometric functions. You are familiar with the real limit

$$\lim_{x\to\infty} \exp(x) = \infty.$$

However, such a result is not true when we replace the real x by a complex z. In fact, given any complex number $w \neq 0$, we have seen that there exists z such that $\exp(z) = w$. But then $\exp(z + 2n\pi\imath) = w$ for all n. Hence we can get z' having arbitrarily large modulus such that $\exp(z') = w$. As a consequence, it follows that $\lim_{z\longrightarrow\infty} \exp(z)$ does not exist. Using the formula for sin and cos in terms of exp, it can be easily shown that sin and cos are both surjective mappings of $\mathbb{C}$ onto $\mathbb{C}$. In particular, remember that they are not bounded unlike their real counter parts.

Example 2.4.3 Logarithm: We would like to define the *logarithm* as the inverse of exponential. However, as we have observed, unlike in the real case, the complex exponential function e^z is not one-one, and hence its inverse is going to be a multi-valued function, or rather a set valued function. Thus indeed, for any $z \neq 0$ (this is needed!) all $w \in \mathbb{C}$ satisfying the equation

$$e^w = z$$

put together in a set is defined as $\log z$ (or $\ln z$). Putting $w = u + \imath v, \quad z = re^{\imath\theta}$, we see that $e^u = r$, and $v = \theta$. Thus $w = \ln z := \ln r + \imath\theta = \ln|z| + \imath \arg z$. Observe that the multi-valuedness of $\ln z$ is a consequence of the same property of $\arg z$.

We have the identity

$$\boxed{\boxed{\ln(z_1 z_2) = \ln z_1 + \ln z_2,}} \tag{2.39}$$

which directly follows from $e^{w_1+w_2} = e^{w_1}e^{w_2}$. Here, we have to interpret this identity 'set-theoretically'. The *principal value* of the logarithm is a single valued function defined by

$$\text{Log } z = \ln|z| + \imath \text{Arg } z.$$

Recall that $-\pi < Arg\, z \leq \pi$. The notation Ln z is also in use. We shall respect both of them. *Caution:* When $z = r$ is a positive real number, $\ln z$ has two meanings! Unless mentioned otherwise one should stick to the older meaning, viz., $\ln r = \text{Ln}\, r$ in that case.

The logarithmic function is all too important to be left as a mere set-valued function. If we restrict the domain suitably, then we see that the 'argument' can be defined continuously. In fact for this to hold, we must be careful that in our domain, which necessarily excludes the origin, we are not able to 'go around' the origin. Thus for instance, if we throw away an entire ray emerging from the origin, from the complex plane, then for each point of the remaining domain a continuous value of the argument can be chosen. This in turn, defines a continuous value of the logarithmic function also. We make a formal definition.

Definition 2.4.1 Given a multi-valued function ϕ, on a domain Ω, by a *branch* of ϕ we mean a specific continuous function $\psi : \Omega \longrightarrow \mathbb{C}$ such that $\psi(z) \in \phi(z)$ for all $z \in \Omega$.

In general for any function f which is not one-one, its inverse is a multi-valued function. Then any continuous function g such that $g \circ f = Id$ over a suitable domain will be called a branch of f^{-1}. In particular, branches of the inverse of the exponential function are called branches of the logarithmic function. Over domains such as $\mathbb{C} \setminus L$ where L is an infinite half-ray from the origin, we easily see that ln has countably infinite number of branches.

The idea of having such a definition is fully justified by the following lemma:

Lemma 2.4.1 Branch lemma: Let Ω_1, Ω_2 be domains in $\mathbb{C}$ and let $f : \Omega_1 \longrightarrow \Omega_2$ be a complex differentiable function, $g : \Omega_2 \longrightarrow \Omega_1$ be a continuous function such that $f \circ g(w) = w, \;\; \forall \;\; w \in \Omega_2$. Suppose $w_0 \in \Omega_2$ is such that $f'(z_0) \neq 0$, where $z_0 = g(w_0)$. Then g is $\mathbb{C}-$differentiable at w_0, with $g'(w_0) = (f'(z_0))^{-1}$.

Proof : Since,

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h} = f'(z_0) \neq 0,$$

it follows that,

$$\lim_{h \to 0} \frac{h}{f(z_0 + h) - f(z_0)} = \frac{1}{f'(z_0)}.$$

Therefore given $\epsilon > 0$ there exists $\delta_1 > 0$, such that,

$$\left| \frac{h}{f(z_0 + h) - f(z_0)} - \frac{1}{f'(z_0)} \right| < \epsilon, \; \forall \; 0 < |h| < \delta_1. \tag{2.40}$$

Now, let $\delta_2 > 0$ be sufficiently small so that $B_{\delta_2}(w_0) \subset \Omega_2$. For $0 < |k| < \delta_2$, put $h := g(w_0 + k) - g(w_0) \neq 0$. By the continuity of g, there exists δ_3 such that $\delta_2 > \delta_3 > 0$, and

$$0 < |k| < \delta_3 \implies |h| = |g(w_0 + k) - g(w_0)| < \delta_1. \tag{2.41}$$

Since $g(w_0) = z_0$, we have, $g(w_0 + k) = z_0 + h$ and hence, $f(z_0 + h) = w_0 + k$. Thus, $k = f(z_0 + h) - f(z_0)$. Therefore, whenever $0 < |k| < \delta_2$, from (2.40) and (2.41), we have,

$$\left| \frac{g(w_0 + k) - g(w_0)}{k} - \frac{1}{f'(g(w_0))} \right| = \left| \frac{z_0 + h - z_0}{f(z_0 + h) - f(z_0)} - \frac{1}{f'(z_0)} \right| < \epsilon.$$

This completes the proof of the lemma. ♠

Remark 2.4.3

1. Observe that as a corollary, we have obtained $\mathbb{C}$-differentiable branches of the logarithmic function. For instance, $l(z) := \ln r + \imath\theta, \quad -\pi < \theta < \pi$, is one such branch defined over the entire of $\mathbb{C}$ minus the negative real axis. The question of the nature of domains on which ln has well defined branches will be discussed later on.

2. The hypothesis that $f'(z_0) \neq 0$ is indeed unnecessary in the above lemma. This stronger version of the above lemma will be proved in Ch.5. In contrast, in the real case, consider the function $x \mapsto x^3$ which defines a continuous bijection of the real line onto itself. Its inverse is also continuous but not differentiable at 0.

Example 2.4.4 Let us find out the derivative of a branch $l(z)$ of the logarithm as a simple application of chain rule. Since, $\exp \circ l = Id$, it follows that $(\exp)'(l(z))l'(z) = 1$. Since $(\exp)' = \exp$, this implies that $zl'(z) = 1$ and hence, $l'(z) = 1/z$.

Remark 2.4.4 Likewise, we could also discuss the 'inverse' of trigonometric functions. They too are multi-valued. However, for continuous choice of a branch over a suitable open set, they will be $\mathbb{C}$-differentiable. For example, any continuous function f such that $\sin(f(z)) = z$ is called a branch of inverse sine function and is denoted by $\arcsin z$. Usually, this notation is reserved for the principal branch i.e., $-\pi < \Re(\arcsin z) \leq \pi$. Due to the periodicity of sine, any two branches of arcsin differ by a a constant $2n\pi, n \in \mathbb{Z}$.

On the other hand, exercises 8-18 in the previous section allow us to define the formal inverse of a convergent power series p with $p'(0) \neq 0$. This inverse power series has a positive radius of convergence and hence defines an analytic function in a small neighborhood of $p(0)$ which is the inverse of the function defined by p. In particular, this can well be invoked to define all the inverse trigonometric functions. In practice however, writing down general formula for the n^{th} term of the inverse power series is not easy. Nor it is easy to determine the exact radius of convergence of such a series. We shall deal with these problems in a function theoretic way.

In any case, we now have a large class of $\mathbb{C}$-differentiable functions –polynomials, exponential function, trigonometric functions. We can also take linear combinations of these and their quotients. We can even consider the 'well chosen' branches of the inverses of these functions and so on. Together, all these constitute what is known as the class of elementary functions. In a latter section we shall again discuss mapping properties of some of them.

Example 2.4.5 Exponents of complex numbers

Recall that defining exponents was somewhat an involved process, even with positive real numbers. Now, we want to deal with this concept with complex numbers. Here the idea is to use the logarithm function which converts multiplication into addition and hence the exponent into multiplication.

For any two complex numbers $z, w \in \mathbb{C} \setminus \{0\}$, define

$$\boxed{\boxed{z^w := e^{w \ln z}.}} \tag{2.42}$$

Observe that on the rhs the term $\ln z$ is a multi-valued function. Therefore, in general, this makes z^w a set of complex numbers rather than a single number. For instance, $2^{1/2}$ is a two element set viz., $\{\sqrt{2}, -\sqrt{2}\}$. Also, wherever we have a branch of $\ln z$ we obtain a $\mathbb{C}$-differentiable function of z. Moreover, it is always a $\mathbb{C}$-differentiable function of the variable w, for a fixed value of $\ln z$.

First, let us take the simplest case, viz., $w = n$ a positive integer. Then irrespective of the value of z, (2.42) gives the single value which is equal to z multiplied with itself n times. For negative integer exponents also, the story is the same. But as soon as w is not an integer, we can no longer say that this is single-valued.

Does this definition follow the familiar laws of exponents:

$$z^{w_1+w_2} = z^{w_1} z^{w_2}; (z_1 z_2)^w = z_1^w z_2^w? \tag{2.43}$$

Yes indeed. The only caution is that these formulae tell you that the two terms on either side of the equality sign are equal as sets. To prove this, use (2.39) repeatedly.

Exercise 2.4

1. Work out the following formulae, where, $z = x + \imath y$:

(a) $\cos z = \cos x \cosh y - \imath \sin x \sinh y$; (b) $\sin z = \sin x \cosh y + \imath \cos x \sinh y$;

(c) $\cosh z = \cosh x \cos y + \imath \sinh x \sin y$; (d) $\sinh z = \sinh x \cos y + \imath \cosh x \sin y$;

(e) $\cosh \imath y = \cos y$; $\sinh \imath y = \imath \sin y$; (f) $\cosh^2 z - \sinh^2 z = 1$.

(g) $\dfrac{1}{\sin z} = \cot z + \tan \dfrac{z}{2}$; (h) $\cot' z + \cot^2 z + 1 = 0$;

2. Prove the **double angle formula:** $2 \cot 2z = \cot z + \cot(z + \pi/2)$.

3. Show that $\cot u \cot v + \cot v \cot w + \cot w \cot u = 1$ if $u + v + w = 0$.

4. Find the limit: $\lim_{y \longrightarrow \infty} \cot \pi(x + \imath y)$.

5. Show that $|\cot \pi(1/2 + \imath y)| < 1$.

6. Compute $\dfrac{d}{dz}(\cosh z)$, $\dfrac{d}{dz}(\sinh z)$.

7. We say a function f is periodic if there exists $w \neq 0$ such that $f(z + w) = f(z)$ for all $z \in \mathbb{C}$. Any such w is called a period of f. If $|w|$ is the smallest amongst all such w then we say that w is *the period* of f. Show that trigonometric and hyperbolic functions are periodic and find their periods.

8. Determine the range of sine and cosine functions. Are they bounded on $\mathbb{C}$?

9. Find the value of e^z for $z = -\frac{\pi\imath}{2}, \frac{3\pi\imath}{4}, \frac{2\pi\imath}{3}$.

10. Find $\ln w$ for $w = e, 2, -1, \imath, 1 + 2\imath$.

11. Show that (i) $\overline{e^z} = e^{\bar{z}}$; (ii) $\overline{\ln z} = \ln \bar{z}$; (iii) $\overline{z^w} = \bar{z}^{\bar{w}}$.

12. Write down all the branches of $\ln z$ on the domain $\mathbb{C} \setminus \{r\imath \ : \ r \geq 0\}$.

13. Express e^{e^z} in the form $u + \imath v$, where, $z = x + \imath y$.

14. Compute z^z in terms of x and y where $z = x + \imath y$. Also, write down explicitly, the value sets for $\imath^\imath$ and $\imath^{-\imath}$.

15. Prove $|\cos z| \geq \sinh |y|$, where $z = x + iy$.

16. Find all values of z for which (a) $\cos z$ (b) $\sin z$ are real.

17. Show that all solutions of (a) $\cos z = 0$ (b) $\sin z = 0$ are real.

18. Solve (i) $\ln z = \frac{1}{2}\pi i$; (ii) $\cos z = 1$; (iii) $\sin z = \imath$.

19. Express the principal values of the following in the form $x + \imath y$.

    (i) $(1+i)^i$; (ii) 3^{3-i}.

20. Show that :

    (i) $\sin^{-1} z = -i \ln(iz \pm \sqrt{1-z^2})$;

    (ii) $\cos^{-1} z = -i \ln(z \pm \sqrt{z^2-1})$;

    (iii) $\tan^{-1} z = \frac{i}{2} \ln\left(\frac{i+z}{i-z}\right)$;

    (iv) $\sinh^{-1} z = \ln(z \pm \sqrt{z^2+1})$;

    (v) $\cosh^{-1} z = \ln(z \pm \sqrt{z^2-1})$;

    (vi) $\tanh^{-1} z = \frac{1}{2} \ln\left(\frac{1+z}{1-z}\right)$

2.5 Miscellaneous Exercises to Ch. 2

1. Find (a) $\Re(e^{\imath z^2})$; (b) $\Im(e^{\sin z})$.

2. Find all the values of $\sqrt{3+4\imath} + \sqrt{3-4\imath}$.

3. Find all complex numbers z such that $z^n = \bar{z}, n \in \mathbb{Z}$.

4. Describe geometrically the region defined by $|e^{-\imath z}| < 1$.

5. For any $n \geq 2$, express $\cos^n x$ as a linear combination of $1, \cos x, \cos 2x, \ldots, \cos nx$.

6. Show that $\displaystyle\int_0^{\pi/2} \cos^{2n} x \, dx = \frac{\pi}{2} \frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots 2n}$.

7. If $z = re^{\imath\theta}, z \neq 1$, show that $\Re[\ln(z-1)] = \dfrac{1}{2} \ln(1 - 2r\cos\theta + r^2)$.

8. Find all values of (a) $\imath^\imath$; (b) $\ln(\imath^{1/2})$; (c) $\ln[\ln(\cos\theta + \imath \sin\theta)]$, $0 \leq \theta < 2\pi$.

9. Find all solutions of $z^2 = \left(\cos \dfrac{2\pi}{3} + \imath \sin \dfrac{2\pi}{3}\right)^3$.

10. Determine the range of cosec, sec and cot.

11. Find all maximal open infinite horizontal strips on which e^z is injective. What is the image of these strips?

12. Find the derivatives of (i) 2^z (ii) $z^\imath$ (iii) z^z.

13. Derive the identity $\operatorname{Exp}(a+b) = (\operatorname{Exp} a)(\operatorname{Exp} b)$ by using the uniqueness of the solution of an initial value problem.

14. Pin-point the mistake in the following:

$$-1 = \imath \imath = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{1} = 1??$$

15. **Mystery of the argument resolved *** Here is a rigorous way to treat the notion of angle or argument of a non zero complex number. The results of section 2.1-2.3 and that of section 2.4 up to (2.27) do not depend on the intuitive notion of angle that we have used elsewhere so far. So we shall take up this study here from that point onwards. Prove the following sequence of exercises to arrive at a rigorous definition of the argument.

    (i) Define $\sin z$ and $\cos z$ by the formula (2.32) and verify that they are entire functions satisfying:

    $$(\sin z)' = \cos z;\ \ (\cos z)' = -\sin z;\ \ e^{\imath y} = \cos y + \imath \sin y;\ \ \cos^2 z + \sin^2 z = 1.$$

    (ii) $\cos 0 = 1, \ \sin 0 = 0$.
    (iii) There exists $y_0 > 0$ such that $\cos y > 0, \ \ \forall \ \ y \in [0, y_0]$.
    (iv) $\sin y_0 > 0$.
    (v) If $y_1 > y_0$ is such that $\cos y > 0$ for all $y \in [y_0, y_1]$ then prove that $y_1 < y_0 + \frac{\cos y_0}{\sin y_0}$.
    (vi) Conclude that there exists $\phi \in [y_0, y_0 + \frac{\cos y_0}{\sin y_0}]$ such that $\cos \phi = 0$.
    (vi) There exists a least positive ϕ such that $\cos \phi = 0$ and we shall denote it by $\pi/2$. (This is the definition of π.)
    (vii) In the interval $[0, \pi/2]$, $\sin y$ is strictly increasing from 0 to 1 and $\cos y$ is strictly decreasing from 1 to 0.
    (viii) Given $0 \le u \le 1$, there is a unique $\theta \in [0, \pi/2]$ such that $\cos \theta = u$.
    (ix) Given u, v such that $0 \le u, v \le 1$ and $u^2 + v^2 = 1$, there is a unique $\theta \in [0, \pi/2]$ such that $e^{\imath \theta} = u + \imath v$.
    (x) Given any $z \in \mathbb{C}$ with $|z| = 1$, there is a unique $\theta \in [0, 2\pi)$ such that $e^{\imath \theta} = z$. We can now define the principal value of argument by $\operatorname{Arg} z = \theta$.

Chapter 3

Conformality

3.1 Cauchy–Riemann Equations

Let U be an open subset of $\mathbb{C}$, $z_0 = (x_0, y_0) \in U$ and $f : U \longrightarrow \mathbb{C}$ be a function which is complex differentiable at z_0. Thus

$$f'(z_0) := \frac{df}{dz}(z_0) := \lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}. \tag{3.1}$$

Write $z_0 + h = (x_0 + h_1) + \imath(y_0 + h_2)$. We shall compute the limit (3.1) in two diferent ways. First let us take the limit along the line $(x_0 + h_1, y_0), \quad h_1 \in \mathbb{R}$. This amounts to putting $h_2 = 0$ in (3.1). Then the rhs yields the partial derivative of f w.r.t. x at the point z_0. Therefore we obtain

$$f'(z_0) = f_x(z_0). \tag{3.2}$$

Similarly, putting $h_1 = 0$ in (3.1), we get

$$f'(z_0) = \lim_{h_2 \to 0} \frac{f(x_0, y_0 + h_2) - f(x_0, y_0)}{\imath h_2} = \frac{f_y(x_0, y_0)}{\imath} = -\imath f_y(z_0). \tag{3.3}$$

Combining (3.2) with (3.3) we have,

$$\boxed{\boxed{f_x + \imath f_y = 0.}} \tag{3.4}$$

Putting $f = u + \imath v$ where u, v are real valued functions, this can be written in a more elaborate fashion:

$$\boxed{\boxed{u_x(x_0, y_0) = v_y(x_0, y_0); \quad u_y(x_0, y_0) = -v_x(x_0, y_0).}} \tag{3.5}$$

These are called Cauchy–Riemann[1](CR)-equations. Observe that we now have four different expressions for f'

$$\boxed{f' = u_x + \imath v_x = v_y - \imath u_y = u_x - \imath u_y = v_y + \imath v_x.} \tag{3.6}$$

Similarly,

$$\boxed{|f'(z_0)|^2 = u_x^2 + v_x^2 = u_y^2 + v_y^2 = u_x^2 + u_y^2 = v_x^2 + v_y^2 = u_x v_y - u_y v_x.} \tag{3.7}$$

The last expression above, which is the determinant of the matrix

$$\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \tag{3.8}$$

is called the *Jacobin* of the mapping $f = (u, v)$, with respect to the variables (x, y) and is denoted by

$$\boxed{J_{(x,y)}(u, v) := u_x v_y - u_y v_x.} \tag{3.9}$$

Example 3.1.1 CR equations in polar coordinates :

Take a point other than the origin. (At the origin polar coordinate is singular.) Say, $z_0 = (x_0, y_0) \neq (0, 0)$ and let $f = u + \imath v$. We claim that equations

$$\boxed{r u_r = v_\theta; \quad r v_r = -u_\theta.} \tag{3.10}$$

are equivalent to CR-equations and obtain the formula:

$$\boxed{f'(z_0) = e^{-i\theta_0}(u_r + i v_r) = -\frac{i}{z_0}(u_\theta + i v_\theta).} \tag{3.11}$$

Since $x = r\cos\theta; y = r\sin\theta$, by the chain rule, we have

$$\begin{aligned}
x_r &= \cos\theta; & x_\theta &= -r\sin\theta; \\
y_r &= \sin\theta; & y_\theta &= r\cos\theta. \\
u_r &= u_x\cos\theta + u_y\sin\theta; & u_\theta &= -u_x r\sin\theta + u_y r\cos\theta; \\
v_r &= v_x\cos\theta + v_y\sin\theta; & v_\theta &= -v_x r\sin\theta + v_y r\cos\theta.
\end{aligned}$$

[1]Bernhard Riemann(1826-1866) a German mathematician. His ideas concerning geometry of space had a profound effect on the development of modern theoretical physics. He clarified the notion of integral by defining what we now call the Riemann integral. He is one of the trios whose work has immense influence in complex analysis, the other two being Cauchy and Weierstrass.

The last four identities can be expressed in the matrix form:

$$r\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}u_x & v_y\\ u_y & -v_x\end{pmatrix} = \begin{pmatrix}ru_r & v_\theta\\ u_\theta & -rv_r\end{pmatrix}.$$

Since $r \neq 0$ and since the left-most matrix defines an invertible transformation, it follows that column vectors of the second matrix are equal to each other (CR equations in Cartesian coordinates hold) iff column vectors in the last matrix are equal to each other (CR equations in polar coordinates.)

Rewriting the above matrix equation in the form

$$\begin{pmatrix}u_x & v_y\\ u_y & -v_x\end{pmatrix} = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}u_r & v_\theta/r\\ u_\theta/r & -v_r\end{pmatrix}$$

and substituting in $f'(z) = u_x - \imath u_y$, gives

$$\begin{aligned} f'(z) &= u_r\cos\theta - \frac{u_\theta}{r}\sin\theta - \imath\left(u_r\sin\theta + \frac{u_\theta}{r}\cos\theta\right)\\ &= u_r\cos\theta + v_r\sin\theta - \imath u_r\sin\theta + \imath v_r\cos\theta\\ &= e^{-\imath\theta}(u_r + \imath v_r)\\ &= \frac{e^{-\imath\theta}}{r}(v_\theta - \imath u_\theta) = \frac{-\imath}{z}(u_\theta + \imath v_\theta). \end{aligned}$$

Remark 3.1.1 A simple minded application of CR-equations is that it helps us to detect easily when a function is **not** $\mathbb{C}$-differentiable. For example, $\Re(z), \Im(z)$ etc. are not complex differentiable anywhere. The function $z \mapsto |z|^2$ is not complex differentiable for any point $z \neq 0$. However, it satisfies the CR-equations at 0. That of course does not mean that it is $\mathbb{C}$-differentiable at 0. (See the exercise below.) In the example below, we give a slightly different kind of application of CR equations.

Example 3.1.2 Let Ω be a domain and $f : \Omega \to \mathbb{C}$ be a $\mathbb{C}-$differentiable function
(i) which takes only real values OR
(ii) which takes only values of modulus 1. Then we claim that f is a constant.

In (i), it happens that $v = 0$ and hence $v_x = 0 = v_y$. Therefore, from CR equations, $u_x = 0 = u_y$. Therefore, also $v_x = 0 = v_y$.

In (ii), we have $u^2 + v^2 = 1$. Hence, $uu_x + vv_x = 0 = uu_y + vv_y$. Since $(u, v) \neq (0, 0)$ this means $u_xv_y - u_yv_x = 0$. CR equations now give you $u_x^2 + u_y^2 = 0 = v_x^2 + v_y^2$. Therefore, $u_x = u_y = v_x = v_y = 0$.

Thus, in either case, u and v are constants and hence f is a constant.

Remark 3.1.2 In order to understand the full significance of CR-equations, we must know a little more about calculus of two real variables. In the next section, we recall some basic results in real multi-variable calculus and then study the close relationship between complex differentiation and the real total differentiation. You may choose to skip this section and come back to it if necessary while reading the section after that. However, try all the following exercises before going further. If you have difficulty in solving any of them, then perhaps you must read the next section thoroughly.

Exercise 3.1

1. Use CR equations to show that $z \longrightarrow \Re(z), \quad z \longrightarrow \Im(z)$ are not complex differentiable anywhere.

2. Use polar coordinates to show that $z \mapsto |z|^2$ is complex-differentiable at 0. What about the function $z \mapsto |z|$?.

3. Write down all possible expressions for the Cauchy derivative of a complex function $f = u + \imath v$ in terms of partial derivatives of u and v.

4. Verify that the functions $\Re z$, $\Im z$, and $\bar{z}$ do not satisfy CR equations.

5. Show that the function $f(z) = z\Re z$ is complex differentiable only at $z = 0$ and find $f'(0)$. How about $z\Im z$ and $z|z|$?

6. Determine points at which the following functions are complex differentiable. (i) $f(z) = xy + iy$ (ii) $f(z) = e^y e^{ix}$.

7. Let f be a $\mathbb{C}-$differentiable function on an open disc such that its image is contained in a line, a circle, a parabola or a hyperbola. Show that f is a constant.

8. Try to generalize the statement above.

3.2 *Review of Calculus of Two Real Variables

We will need some basic notions of the calculus of two real variables. In this section, we recall these concepts to the extent required to understand the later material that we are going to learn. Indeed, we presume that you are already reasonably familiar with the material of this section.

As a warm-up, we illustrate the kind of danger that we may be in while we are trying to relate the calculus of several variables to that of 1-variable, with an example.

Example 3.2.1 Define a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0). \end{cases} \tag{3.12}$$

This function has questionable behavior only at $(0,0)$. It has the property that for each fixed y, it is continuous for all x and for each fixed x it is continuous for all y. However, it is not continuous at $(0,0)$ even if we are ready to redefine its value at $(0,0)$. This is checked by taking limits along the line $y = mx$. For different values of m, we get different limits at $(0,0)$. So, there is no way we can make it continuous at $(0,0)$.

We begin by recalling the increment theorem of 1-variable calculus.

Theorem 3.2.1 *Let $f : (a,b) \longrightarrow \mathbb{R}$ be a function $x_0 \in (a,b)$. Then f is differentiable at x_0 iff there exists an* error function *η defined in some neighborhood $|x - x_0| < \epsilon$ of x_0 and a real number α such that*

$$\boxed{\boxed{f(x_0+h) - f(x_0) = \alpha h + \eta(h)h, \ \text{and } \eta(h) \longrightarrow 0, \ \text{as } h \longrightarrow 0.}} \tag{3.13}$$

Remark 3.2.1 The proof of this is exactly same as that of theorem 2.1.1. Roughly speaking, the condition in the increment theorem tells us that the difference (increment) in the functional value of f at x_0 is $f'(x_0)$ times the difference (increment) h, in the variable x, up to a second order term viz., $h\eta(h)$. That explains why this result is called the increment theorem.

Remark 3.2.2 Derivative as a Linear approximation: Recall that by a (real) linear map $\phi : \mathbb{R}^n \to \mathbb{R}^m$ we mean a map ϕ having the property

$$\phi(\alpha\mathbf{v} + \beta\mathbf{u}) = \alpha\phi(\mathbf{v}) + \beta\phi(\mathbf{u}), \quad \alpha, \beta \in \mathbb{R}, \mathbf{v}, \mathbf{u} \in \mathbb{R}^n. \tag{3.14}$$

Given a linear map $\phi : \mathbb{R} \to \mathbb{R}$, we put $\phi(1) = \lambda$. Then it is easily checked that $\phi(t) = \phi(t.1) = t\phi(1) = \lambda t$ for all $t \in \mathbb{R}$. Thus every real number $\lambda \in \mathbb{R}$ defines a linear map $\phi : \mathbb{R} \to \mathbb{R}$ via multiplication and vice versa. The graph of this linear map $\phi : \mathbb{R} \to \mathbb{R}$ is nothing but a line passing through the origin with slope λ. Of course, other lines with slope λ are given by functions of the form $\phi(t) = \lambda t + \mu$. For a differentiable function $f : (a,b) \to \mathbb{R}$, we have the approximate formula

$$f(x_0+h) \approx f(x_0) + hf'(x_0)$$

We know that $f'(x_0)$ is the slope of the tangent to the graph of the function $y = f(x)$. Since the tangent line represents an approximation of the graph of the function f, we may say that the linear map corresponding to the tangent line represents an approximation to the function f at x_0. Thus we see that the derivative should be thought of as a *linear map* approximating the given function at the given point. This aspect of the differentiability of a 1-variable function is obscured by the over simplification that occurs naturally in 1-variable linear map viz., *'a linear map $\mathbb{R} \longrightarrow \mathbb{R}$ is nothing but the multiplication by a real number and thus can be identified with that real number'.* When we pass to two or more variables, this simplification disappears and thus the true nature of the derivative comes out, as in the following definition. In what follows, for simplicity, we restrict our attention to two variables, though there is no logical gain in it. All the concepts and results that we are going to introduce for two variables hold good for more number of variables also.

Definition 3.2.1 Let U be an open subset of $\mathbb{R}^2$ and let $f : U \longrightarrow \mathbb{R}$ be any function. Let z_0 be any point in U. We say f is (*Frechet* [2]) *differentiable* at z_0 iff there exists a linear map $L : \mathbb{R}^2 \longrightarrow \mathbb{R}$ and a scalar valued error function η defined in a neighborhood of z_0 in U such that

$$\boxed{f(z_0+h)-f(z_0) = L(h)+\|h\|\eta(h);\ \eta(h) \longrightarrow 0 \text{ as } h \longrightarrow 0.} \tag{3.15}$$

Further L is called the *Frechet derivative* or the *total derivative* of f at z_0 and is denoted by $(Df)_{z_0}$. If f is differentiable at each point of U then it is called a *(Frechet) differentiable function.*

As an easy exercise prove the following theorem:

Theorem 3.2.2 *If f is differentiable at a point then it is continuous at that point.*

Theorem 3.2.3 *Let $U,\ f,\ z_0 = (x_0, y_0)$ etc. be as above. Let f be Frechet differentiable at z_0. Then f has its partial derivatives at z_0 and moreover we have*

$$f_x(z_0) = (Df)_{z_0}(1, 0);\ \ f_y(z_0) = (Df)_{z_0}(0, 1).$$

[2]René Maurice Frechet (1878-1973), a French Mathematician, a student of Hadamard, is best known for his contribution toward laying the foundations of general topology and abstract analysis.

Proof: Let $L = (Df)_{z_0}$. Putting $h = (t, 0)$ in (3.15), we obtain,

$$f(z_0 + t) - f(z_0) = L(t, 0) + \eta(t, 0)|t| = tL(1, 0) + \eta(t, 0)|t|. \tag{3.16}$$

Dividing out by t and taking limit as $t \longrightarrow 0$, it follows that $f_x(z_0)$ exists and is equal to $L(1, 0)$. Similarly, we can show that f_y exists and $f_y(z_0) = L(0, 1)$. ♠

Remark 3.2.3

(i) The concept of partial derivative is a special case of a more general concept. Given any unit vector $\mathbf{u}$, we define the *directional derivative* of f in the direction of $\mathbf{u}$ denoted by $D_{\mathbf{u}}f(z_0)$, to be the limit of

$$\lim_{t \longrightarrow 0} \frac{f(z_0 + t\mathbf{u}) - f(z_0)}{t}$$

provided it exists. As above, it can be seen that all the directional derivatives exist if $(Df)_{z_0}$ exists. Moreover, by putting $h = t\mathbf{u}$, in (3.15), dividing out by t and then taking the limit as $t \longrightarrow 0$, it is verified that $D_{\mathbf{u}}f(z_0) = L(\mathbf{u}) = (Df)_{z_0}(\mathbf{u})$,

(ii) It may happen that all the directional derivatives exist and yet the total derivative $(Df)_{z_0}$ may not exist. This can happen even if all this directional derivatives vanish, as seen in the following two examples.

Example 3.2.2 Consider the function

$$f(x, y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

Clearly f is differentiable at all points except perhaps at $(0, 0)$. We shall show that f is not even continuous at $(0, 0)$ and hence cannot be differentiable at $(0, 0)$. However, observe that if you restrict the function to any of the lines through the origin, then it is continuous. This will tell you that if we approach the origin along any of these lines then the limit of the function coincides with the value of the function. In contrast, in the case of 1-variable function, if the left-hand and right-hand limits existed and agreed with the functional value then the function was continuous at that point. Thus, the geometry of the plane is not merely the geometry of all the lines in it.

In order to see that the function is not continuous at the origin, we shall produce various sequences $\{u_n\}$ such that $\lim_{n \longrightarrow \infty} u_n = (0, 0)$ and $\lim_{n \to \infty} f(u_n)$ takes different values. It then follows that, at $(0, 0)$, even a redefinition of f will not make it continuous. So take any real sequence $x_n \longrightarrow 0$, $x_n \neq 0$ and put $y_n = kx_n^2$ for some real k. Put

$u_n = (x_n, y_n)$. Then $u_n \longrightarrow (0,0)$ and $f(u_n) = k/(1+k^2)$. Therefore, $\lim_{n\longrightarrow\infty} f(u_n) = k/(1+k^2)$. Thus for different values of k we get different values of this limit as required.

On the other hand, let $\mathbf{u} = (a,b)$ be a unit vector. If b is zero then clearly the partial derivative of f in the direction of $\mathbf{u}$ (it is f_x) is zero since the function is identically zero on the x axis. For $b \neq 0$, we have $F_{\mathbf{u}}(t) = a^2bt/(t^2a^4 + b^2)$ for all t. It follows that $D_{\mathbf{u}}f(0,0) = F'_{\mathbf{u}}(0) = a^2b/b^2 = a^2/b$. Thus all the directional derivatives exist. Also, for your own satisfaction check that the partial derivatives are not continuous at $(0,0)$.

Example 3.2.3 We can improve upon the above example 3.2.2, as follows. Take $g(x,y) = \sqrt{x^2+y^2}f(x,y)$, where f is given as in example 3.2.2. Then the function g is continuous also at $(0,0)$ and has all the directional derivatives vanish at $(0,0)$. That means that the graph of this function has the xy-plane as a plane of tangent lines at the point $(0,0,0)$. We are tempted to award such 'nice' geometric behavior of the function and admit it to be 'differentiable' at $(0,0)$. Alas! It is not differentiable at $(0,0)$, in the definition that we have adopted. For

$$\frac{g(x,y) - g(0,0)}{\|(x,y)\|} = f(x,y)$$

has no limit at $(0,0)$.

We hope that the above two examples illustrate the subtlety of the situation in the following theorem, which is a result in the positive direction.

Theorem 3.2.4 *Let U be an open set in $\mathbb{R}^2$ and $f : U \to \mathbb{R}$ be a function having partial derivatives which are continuous at (x_0, y_0). Then f is Frechet differentiable at (x_0, y_0).*

Proof: By Mean Value theorem of 1-variable calculus, there exist $0 \le t, s \le 1$ such that
$f(x_0+h, y_0+k) - f(x_0+h, y_0) = kf_y(x_0+h, y_0+sk)$;
$f(x_0+h, y_0) - f(x_0, y_0) = hf_x(x_0+th, y_0)$. (Of course t, s depend on h, k.) Therefore,
$|f(x_0+h, y_0+k) - f(x_0, y_0) - hf_x(x_0, y_0) - kf_y(x_0, y_0)|$
$\le |h||f_x(x_0+th, y_0) - f_x(x_0, y_0)| + |k||f_y(x_0+h, y_0+sk) - f_y(x_0, y_0)|$.
By continuity of f_x, f_y at (x_0, y_0), given $\epsilon > 0$ we can choose h, k sufficiently small so that $|f_x(x_0+th, y_0) - f_x(x_0, y_0)| < \epsilon$; and $|f_y(x_0+h, y_0+sk) - f_y(x_0, y_0)| < \epsilon$. The conclusion follows. ♠

Remark 3.2.4
(i) Given a linear map $\phi : \mathbb{R}^2 \to \mathbb{R}$ let us write $\phi(1,0) = \alpha$ and $\phi(0,1) = \beta$. It then

follows that ϕ is completely determined the vector (α, β) by the dot product formula:

$$\phi(t,s) = t\alpha + s\beta = (t,s)\cdot(\alpha,\beta) \tag{3.17}$$

Thus we may identify the space of all linear maps $\phi : \mathbb{R}^2 \to \mathbb{R}$ with $\mathbb{R}^2$ itself! (This space is called the linear dual of $\mathbb{R}^2$.)
(ii) In the case when ϕ is the derivative $\phi = Df_{\mathbf{x}}$ of f at a point $\mathbf{x}$, the vector (α, β) is nothing but the vector $(grad\, f)_{\mathbf{x}} = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})_{\mathbf{x}}$. As we have seen earlier, it is possible that $grad\, f$ exists even though it may happen that Df does not exist. A slight variation of the theorem 3.2.4 above can now be stated. The proof is left to you as a simple exercise.

Theorem 3.2.5 *Let U be an open subset of $\mathbb{R}^2$ and $f : U \longrightarrow \mathbb{R}$ be a function. Then f is Frechet differentiable in U and the function $Df : U \longrightarrow \mathbb{R}^2$ is continuous iff both the partial derivatives of f exist on U and are continuous on U.*

Definition 3.2.2 Given a function $f : U \to \mathbb{R}$, we say f is *continuously differentiable* on U if Df is defined in U and all the first order partial derivatives are continuous. In this case, we also say that f is of *class $\mathcal{C}^1$ in U*. Inductively, a function f on U is said to be of *class $\mathcal{C}^r$ in U* if all the partial derivatives of f of order r exist and are continuous in U. Finally, a function which is of class $\mathcal{C}^r$ for all $r > 0$ is said to be of class $\mathcal{C}^\infty$.

Remark 3.2.5 All the standard properties of the derivatives of a function of one variable such as for sums and scalar multiples etc. hold here also with obvious modifications wherever necessary. For instance, if f and g are real valued differentiable functions then their product is differentiable and we have

$$D(fg)_{z_0} = f(z_0)D(g)_{z_0} + g(z_0)D(f)_{z_0}.$$

Caution: Note that Mean Value Theorem is one result which really needs modification form 1-variable to several variables.

We now generalize the above results to the case of vector valued functions. All that we have to do is to apply the corresponding propety coordinatewise to all coordinate functions.

Definition 3.2.3 Let $f : U \longrightarrow \mathbb{R}^m$ be a function and z_0 be a point of the open set $U \subset \mathbb{R}^n$. We say f is differentiable at z_0 if there exists a linear map $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and an error function $\eta : B_r(0) \longrightarrow \mathbb{R}^n$ such that for $h \in B_r(0)$, we have,

$$\boxed{f(z_0 + h) - f(z_0) = L(h) + \|h\|\eta(h);\;\; \lim_{h\to 0}\eta(h) = 0.} \tag{3.18}$$

In this case, We write $D(f)_{z_0} = L$.

Here n, m could be any positive integers. For our purpose, you may just take them ≤ 2.

Let us recall a simple result form linear algebra:

Lemma 3.2.1 Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then there exists $M > 0$ such that for all $h \in \mathbb{R}^n$ we have $\|L(h)\| \leq M\|h\|$.

Proof: Let $\mathbf{e}_i$ denote the standard basis vectors in $\mathbb{R}^k$. Write $L(\mathbf{e}_j) = \sum_{i=1}^m \lambda_{ij}\mathbf{e}_i,\ 1 \leq j \leq n$. Put $\lambda := max\{|\lambda_{ij}|\}$. Then for $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ we have

$$\|L(h)\|^2 = \sum_i \left(\sum_j \lambda_{ij} h_j\right)^2 \leq m\lambda\|h\|^2$$

and so we can choose $M > \sqrt{m\lambda}$. ♠

Theorem 3.2.6 Chain Rule *Let $f : U \longrightarrow V$ and $g : V \longrightarrow \mathbb{R}$ be such that f is differentiable at $z_0 \in U$ and g is differentiable at $w_0 = f(z_0) \in V$. Then $g \circ f$ is differentiable at z_0 and we have,*

$$D(g \circ f)_{z_0} = D(g)_{w_0} \circ D(f)_{z_0}.$$

Proof: We have linear maps L_1, L_2 and error functions η_1, η_2 such that

$$f(z_0 + h) - f(z_0) = L_1(h) + \|h\|\eta_1(h); \quad g(w_0 + k) - g(w_0) = L_2(k) + \|k\|\eta_2(k),$$

where $\eta_1(h) \longrightarrow 0$ as $h \longrightarrow 0$ and $\eta_2(k) \longrightarrow 0$ as $k \longrightarrow 0$. Note that $f(z_0+h)-f(z_0) \longrightarrow 0$ as $h \longrightarrow 0$. Therefore, we can substitute $k = f(z_0 + h) - f(z_0)$ in the second equation. This gives,

$$\begin{aligned} g \circ f(z_0 + h) - g \circ f(z_0) &= L_2(L_1(h) + \|h\|\eta_1(h)) + \|h\| \left(\frac{||k||}{||h||}\eta_2(k)\right) \\ &= L_2 \circ L_1(h) + \|h\| \left(L_2(\eta_1(h)) + \frac{\|k\|}{\|h\|}\eta_2(k)\right) \end{aligned}$$

Observe that $\eta_1(h) \to 0$ as $h \to 0$ and hence $\eta_1(h)$ is bounded in a neighbourhood of 0. By the lemma above $\frac{L_1(h)}{||h||}$ is also bounded. This implies that

$$\frac{\|k\|}{\|h\|} \leq \frac{\|L_1(h)\|}{\|h\|} + \|\eta_1(h)\|$$

is bounded. Therefore, if we take $\eta(h) = L_2(\eta_1(h)) + \frac{\|k\|}{\|h\|}\eta_2(k)$, it follows that $\eta(h) \longrightarrow 0$ as $h \longrightarrow 0$. The result follows. ♠

As an easy consequence we have:

Theorem 3.2.7 *Let U be a convex open subset of $\mathbb{R}^2$ and $f : U \longrightarrow \mathbb{R}$ be a differentiable function such that $D(f)_z = 0$ for all $z \in U$. Then $f(z) = c$, a constant, on U.*

Proof: Fix a point $z_0 \in U$. Now for any point $z \in U$ consider the map $g : [0,1] \longrightarrow U$ given by $g(t) = (1-t)z_0 + tz$. By chain rule the composite map $h := f \circ g : [0,1] \longrightarrow \mathbb{R}$ is differentiable and its derivative vanishes everywhere. By 1-variable calculus, (Lagrange's Mean Value theorem), applied to each component of $h = (h_1, h_2)$, it follows that h is a constant on $[0,1]$. In particular, $h(1) = h(0)$. But $f(z) = h(1) = h(0) = f(z_0)$. ♠

Remark 3.2.6 Observe that the projection maps are differentiable. Therefore, it follows that if $f = (f_1, f_2)$ is differentiable then each co-ordinate function f_j is also so. It is not difficult to see that the converse is also true. The derivatives $D(f_1)$ and $D(f_2)$ can be treated as row vectors and by writing them one below the other, we get a 2×2 matrix $D(f)$. With this notation, the chain rule can be stated in terms of matrix multiplication.

Having identified a linear map $L : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ with a 2×2 real matrix, we see that $D(f)$ is a function from U to $M(2;\mathbb{R})$. The latter space can be identified with the Euclidean space $\mathbb{R}^4$. We can then see that $D(f) : U \longrightarrow \mathbb{R}^4$ is continuous iff the partial derivatives of f_1, f_2 are continuous. The map $f : U \longrightarrow \mathbb{R}^2$ is called a map of class $\mathcal{C}^1$ on U if it is differentiable and the derivative $D(f)$ is continuous. What is then the meaning of $D(f) : U \longrightarrow \mathbb{R}^4$ is differentiable? Going by the above principle, we see that this is the same as saying that all the four component functions of $D(f)$ should be differentiable. The derivative of $D(f)$ is actually a function $D^2(f) : U \longrightarrow \mathbb{R}^8$. Components of this are nothing but the second order partial derivatives of the components of f. Thus for any positive integer k, we define f to be of class $\mathcal{C}^k$ on U if all its k-th order partial derivatives exist and are continuous on U. If f is of class $\mathcal{C}^k$ for all $k \geq 1$ then it is said to belong to the class $\mathcal{C}^\infty$. Such maps are also called smooth maps.

All this can be easily generalized to functions from subsets of $\mathbb{R}^n$ to $\mathbb{R}^m$ for any positive integers m, n.

Exercise 3.2

1. Suppose $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ are continuous functions. Show that each of the following functions on $\mathbb{R}^2$ are continuous.
   (i) $(x,y) \mapsto f(x) + g(y)$; (ii) $(x,y) \mapsto f(x)g(y)$;
   (iii) $(x,y) \mapsto max\{f(x), g(y)\}$; (iv) $(x,y) \mapsto min\{f(x), g(y)\}$.

2. Use the above exercise, if necessary to show that $f(x, y) = x + y$ and $g(x, y) = xy$ are continuous functions on $\mathbb{R}^2$. Deduce that every polynomial function in two variables is continuous. Can you generalize this?

3. Express the definition of $\lim\limits_{(x,y)\to(0,0)} f(x, y)$ in terms of polar coordinates and use it to analyze limit of the following functions:
(i) $f(x, y) = \dfrac{x^3 - xy^2}{x^2 + y^2}$; (ii) $g(x, y) = \tan^{-1}\left(\dfrac{|x| + |y|}{x^2 + y^2}\right)$;
(iii) $h(x, y) = \dfrac{y^2}{x^2 + y^2}$.

4. Examine the following functions for continuity at $(0, 0)$. The expressions below give the value of the function at $(x, y) \neq (0, 0)$. At $(0, 0)$ you are free to take any value you like.

$$\begin{array}{ll}
\text{(i) } \dfrac{x^3 y}{x^2 - y^2}; \quad \text{(ii) } \dfrac{x^2 y}{x^2 + y^2}; & \text{(iii) } xy\dfrac{x^2 - y^2}{x^2 + y^2}; \\
\text{(iv) } |[|x| - |y|]| - |x| - |y|; & \text{(v) } \dfrac{\sin^2(x + y)}{|x| + |y|}.
\end{array}$$

5. Examine each of the following functions for continuity.
(i) $f(x, y) = \begin{cases} \dfrac{y}{|y|}\sqrt{x^2 + y^2}, & y \neq 0, \\ 0, & y = 0. \end{cases}$
(ii) $g(x, y) = \begin{cases} x \sin \dfrac{1}{x} + y \sin \dfrac{1}{y}, & x \neq 0, y \neq 0; \\ x \sin \frac{1}{x}, & x \neq 0, y = 0; \\ y \sin \frac{1}{y}, & x = 0, y \neq 0; \\ 0, & x = 0, y = 0. \end{cases}$

6. Let $f : B_r(0) \to \mathbb{R}$ be some function where $B_r(0)$ is the open disc of radius r and centre 0 in $\mathbb{R}^2$. Assume that the two limits

$$l(y) = \lim_{x\to 0} f(x, y), \qquad r(x) = \lim_{y\to 0} f(x, y) \tag{3.19}$$

exist for all sufficiently small y and for all sufficiently small x respectively. Assume further that the limit $\lim\limits_{(x,y)\to(0,0)} f(x, y) = L$ also exists. Then show that the iterated limits

$$\lim_{y\to 0}[\lim_{x\to 0} f(x, y)], \qquad \lim_{x\to 0}[\lim_{y\to 0} f(x, y)] \tag{3.20}$$

both exist and are equal to L.

7. Put $f(x,y) = \dfrac{x-y}{x+y}$, for $(x,y) \neq (0,0)$. Show that the two iterated limits (3.20) exist but are not equal. Conclude that the limit $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist.

8. Put $f(x,y) = \dfrac{x^2y^2}{x^2y^2 + (x-y)^2}$, $(x,y) \neq (0,0)$. Show that the iterated limits (3.20) both exist at $(0,0)$. Compute them. Show that the $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

9. Consider the function

$$f(x,y) = \begin{cases} \dfrac{x^3}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

(a) Show that f is continuous and all the directional derivatives f of exist and are bounded.
(b) For any $\mathcal{C}^1$ mapping $g : \mathbb{R} \to \mathbb{R}^2$ show that $f \circ g$ is a $\mathcal{C}^1$- mapping.
(c) Yet f is not differentiable at $(0,0)$. [Hint: Use polar coordinates.]

3.3 *Cauchy Derivative (Vs) Frechet Derivative

Partial derivatives play a key role in the comparison study of Cauchy derivative and Frechet derivative. We have seen that existence of either of them implies the existence of partial derivatives. Moreover, in the former case, the partial derivatives satisfy the CR-equations. Thus, even if Df exists, if CR equations are not satisfied then f' does not exist. Using this we can give plenty of examples of Frechet differentiable functions which are not $\mathbb{C}$-differentiable. As we have already seen, the geometry of the plane is responsible for making the total derivative somewhat subtler in comparison with the derivative in the case of one 1-variable function. What additional basic structure is then responsible for the difference in Cauchy differentiation and Frechet differentiation? An answer to this question is in the following theorem:

Theorem 3.3.1 *Let $f : U \longrightarrow \mathbb{C}$ be a continuous function, $f = u + \imath v$, $z_0 = x_0 + \imath y_0$ be a point of U. Then f is $\mathbb{C}$-differentiable at z_0 iff considered as a vector valued function of two real variables, f is (Frechet) differentiable at z_0 and its derivative $(Df)_{z_0} : \mathbb{C} \longrightarrow \mathbb{C}$ is a complex linear map. In that case, we also have $f'(z_0) = (Df)_{z_0}$.*

Proof: Recall that a map $\phi : V \longrightarrow W$ of complex vector spaces is complex linear iff $\phi(\alpha v + \beta w) = \alpha\phi(v) + \beta\phi(w)$ for any $\alpha \in \mathbb{C}$ and $v, w \in V$. Let us first consider a purely algebraic problem: Treating $\mathbb{C}$ as a 2-dimensional real vector space, consider a real linear map $T : \mathbb{C} \longrightarrow \mathbb{C}$ given by the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3.21}$$

When is it a complex linear map? We see that, if T is complex linear, then $T(\imath) = \imath T(1)$ and hence, $b + \imath d = T(\imath) = \imath T(1) = \imath(a + \imath c)$. Therefore, $b = -c$ and $a = d$. Conversely, it is easily seen that this condition is enough to ensure the complex linearity of T.

Coming to the proof of the theorem, suppose that f is $\mathbb{C}$-differentiable at z_0. Then as already seen the partial derivatives exist and satisfy the Cauchy–Riemann equations. So, the 2×2 matrix (3.21) defines a complex linear map from $\mathbb{C}$ to $\mathbb{C}$. It remains to see that f is real differentiable at z_0, for then, automatically the derivative will be equal to the matrix (3.21) above. For this, we directly appeal to the increment theorem: We have,

$$f(z_0 + h) - f(z_0) = (\alpha + \imath\beta)h + h\eta(h),$$

where, $\alpha = u_x = v_y,\ \beta = v_x = -u_y$. Put $\phi(h) = \dfrac{h}{\|h\|}\eta(h)$. Then,

$$\lim_{h\to 0} \|\phi(h)\| = \lim_{h\to 0} \|\eta(h)\| = 0.$$

Also, the multiplication map $h \mapsto (\alpha + \beta\imath)h$ can be viewed as a real linear map acting on the 2-vector h, it follows that f is Frechet differentiable, with the derivative Df given by (3.21).

Conversely, assume that $Df_{z_0} = T$ exists and is complex linear. By (3.18), we have,

$$f(z_0 + h) - f(z_0) = T(h) + |h|\eta(h), \quad \lim_{h\to 0} \eta(h) = 0.$$

If T is given by the matrix (3.21), then we know that $b = -c, d = a$ and hence $T(h) = \lambda h$ where $\lambda = a + \imath c$. Therefore

$$f(z_0 + h) - f(z_0) = \lambda h + |h|\eta(h), \quad \lim_{h\to 0} \eta(h) = 0$$

which is equivalent to (2.5). By Increment theorem 2.1.1, it follows that f ic $\mathbb{C}$-differentiable at z_0 with $f'(z_0) = \lambda = a + \imath c = u_x + \imath v_x$. ♠

There are certain useful partial results that relate the two notions of differentiability. Let us begin with a definition:

Definition 3.3.1 Let U be an open subset of $\mathbb{C}$ and $f : U \to \mathbb{C}$ be a continuous function with continuous partial derivatives of the first order. We say f is holomorphic on U if it satisfies the CR equations throughout U.

Theorem 3.3.2 *Let f be a continuous complex valued function of a complex variable defined on an open subset U, possessing continuous partial derivatives. Then f is holomorphic on U iff f is complex differentiable in U.*

Proof: The 'if' part has been proved already. The 'only if' part is the consequence of theorems 3.2.5 and 3.3.1 along with the observation that CR-equations are equivalent to say that the Frechet derivative is complex linear.

Remark 3.3.1 The continuity hypothesis on the partial derivatives can be removed in the above theorem. Then the proof of the **'only if'** part is precisely as above. However, you have to wait for the proof of the **'if'** part till chapter 4 where we shall actually prove that a complex differentiable function is analytic. Indeed, perhaps the strongest result in this direction is what is known as **Looman-Menchoff Theorem** the proof of which involves ideas that are beyond the theme of this course. Interested reader can look in [N].

Theorem 3.3.3 Looman-Menchoff *Let U be an open subset of $\mathbb{C}$ and $f : U \longrightarrow \mathbb{C}$ be a continuous function, $f = u + \imath v$. Suppose the partial derivatives of u, v exist and satisfy Cauchy-Riemann equations throughout U. Then f is complex differentiable in U.*

Remark 3.3.2 There are many functions that are complex differentiable at a point but not so at any other points in a neighborhood (see the exercises below). As far as the differentiation theory is concerned such functions are not of much use to us. We would like to concentrate on those functions which are differentiable in some non empty open subset of $\mathbb{C}$. Clearly holomorphic functions come under this class. However, checking CR-equations is the last thing we may do to verify whether a given function is complex differentiable or not. On the other hand, we know that analytic functions are indeed holomorphic and hence complex differentiable. Right now convergent power series are our best source for complex differentiable functions. Later, in chapter 4, we shall see that every complex differentiable function in a domain is actually analytic as well and thus, all the three notions coincide. Thus for us, the terminology 'holomorphic' is only a temporary convenience. Nevertheless, we may not ignore it since it is extensively used in the literature.

$$\begin{tikzcd}
& \text{Analyticity} \arrow[dl] & \\
\text{Holomorphicity} \arrow[rr] & & \text{Complex Differentiability} \arrow[ul, dashed]
\end{tikzcd}$$

Notice the dashed arrow. This love-triangle will be completed in chapter 4.

Exercise 3.3

1. Show that $|z|$ is Frechet differentiable everywhere except at $z = 0$. Can you say the same thing about complex differentiability?

2. Show that $z|z|$ is complex differentiable only at $z = 0$ and Frechet differentiable everywhere.

3. Show that $f(x, y) = \sqrt{|xy|}$ is continuous and has partial derivatives which satisfy C-R equation at $(0, 0)$, yet f is not complex differentiable at $(0, 0)$. Does this contradict Looman-Menchoff theorem?

4. Establish the following generalization of Cauchy-Riemann equations. "If $f(z) = u + iv$ is differentiable at a point $z_0 = z_0 + iy_0$ of a domain G, then
$$\frac{\partial u}{\partial s} = \frac{\partial v}{\partial n}, \quad \frac{\partial u}{\partial n} = -\frac{\partial v}{\partial s} \tag{3.22}$$
at (x_0, y_0) where $\frac{\partial}{\partial s}$ and $\frac{\partial}{\partial n}$ denote directional differentiation in two orthogonal directions s and n at (x_0, y_0), such that n is obtained from s by making a counter-clockwise rotation."

5. Let $f : \Omega_1 \to \mathbb{C}$ be a complex differentiable function. Suppose L_1 is a line segment contained in Ω_1 such that $f(L_1)$ is contained in a line segment L_2. Consider the restriction of the f on the segment L_1 viz., the function $g : L_1 \to L_2$ given by $z \mapsto f(z)$ as a function of one real variable. Show that g is differentiable and $g'(z_0) = 0$ iff $f'(z_0) = 0$. Give an example to show that such a statement is not true if we replace '*f is complex differentiable*' by '*f is real (Frechet) differentiable*'.

3.4 *Formal Differentiation and an Application

Guided by the chain rule for differentiation and the following basic algebraic relations

$$z = x + \imath y,\ \bar{z} = x - \imath y,\ x = \frac{z+\bar{z}}{2},\ y = \frac{z-\bar{z}}{2\imath},$$

we introduce the following partial differential operators:

$$\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - \imath\frac{\partial}{\partial y}\right);\ \frac{\partial}{\partial \bar{z}} := \frac{1}{2}\left(\frac{\partial}{\partial x} + \imath\frac{\partial}{\partial y}\right). \tag{3.23}$$

Assume that $f = u + \imath v$ is a continuous function having continuous first order partial derivatives. Then the CR equations (3.4) for f can be expressed in a single equation $\dfrac{\partial f}{\partial \bar{z}} = 0$. Moreover, if f is holomorphic, then we have,

$$\frac{\partial f}{\partial z} = \frac{1}{2}(u_x - \imath u_y) + \frac{\imath}{2}(v_x - \imath v_y) = u_x + \imath v_x = \frac{df}{dz} = f'(z).$$

Observe that $\dfrac{\partial z}{\partial z} = 1$ and $\dfrac{\partial \bar{z}}{\partial \bar{z}} = 1$. Thus, for a holomorphic function f we have

$$\boxed{\frac{\partial f}{\partial \bar{z}} = 0;\ \ \frac{\partial f}{\partial z}(z) = f'(z).} \tag{3.24}$$

Remark 3.4.1 Thus a function having continuous partial derivatives is holomorphic iff $\frac{\partial f}{\partial \bar{z}} = 0$. In this sense, one can say that '*holomorphic functions are independent of* $\bar{z}$.' This seemingly bizarre statement can be made very precise using power series in two variables which is beyond the scope of this book. (Interested reader may have a look into section 3 of chapter 4 in Cartan's book [Car]). One can also study functions which have the property

$$\frac{\partial f}{\partial z} = 0.$$

These are called *anti-holomorphic* functions. Because of the close association of them with holomorphic functions, as seen below, we need not even study them separately.

Lemma 3.4.1 Given a complex valued function f of a complex variable, the following are equivalent:

(a) $f(z)$ is holomorphic at z_0;

(b) $f(\bar{z})$ is anti-holomorphic at $\bar{z}_0$;

(c) $\overline{f}(z) := \overline{f(z)}$ is anti-holomorphic at z_0.

Proof: Put $f(x,y) = u(x,y) + \imath v(x,y)$ and $g(x,y) := f(\bar{z}) =: p(x,y) + \imath q(x,y)$. Then we have, $p(x,y) = u(x,-y)$ and $q(x,y) = v(x,-y)$. Therefore, $p_x(x,-y) = u_x(x,y), p_y(x,-y) = -u_y(x,y), q_x(x,-y) = v_x(x,y), q_y(x,-y) = -v_y(x,y)$. From this it follows that

$$\frac{\partial f}{\partial \bar{z}}(z_0) = \frac{\partial g}{\partial z}(\bar{z}_0).$$

The equivalence of (a) and (b) follows. Verify the equivalence of (a) and (c) in a similar fashion. ♠

Theorem 3.4.1 *Let f and g be complex valued functions such that $f \circ g$ is defined.*
(i) *If one of f, g is holomorphic and the other anti-holomorphic, then $f \circ g$ is anti-holomorphic;*
(ii) *If f and g are both anti-holomorphic then $f \circ g$ is holomorphic.*

Proof: Suppose f is holomorphic and g is anti-holomorphic, then $g(\bar{z})$ is holomorphic and hence $f \circ g(\bar{z})$ is holomorphic. Therefore, $f \circ g$ is anti-holomorphic. Other statements are proved similarly. ♠

Exercise 3.4 Recall that the Laplace operator is defined by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Obviously this can be applied to functions of two variables which possess second order partial derivatives. We shall assume here that all functions possess continuous second order partial derivatives.

1. Establish the following formula: $\nabla^2 u = 4\dfrac{\partial^2}{\partial z \partial \bar{z}} u = 4\dfrac{\partial^2}{\partial \bar{z} \partial z} u.$

2. If f is holomorphic, show that $\nabla^2(\ln(1 + |f|^2)) = \dfrac{4|f'|^2}{(1 + |f|^2)^2}.$

3. Let $f : \Omega_1 \to \Omega_2$ be a holomorphic mapping and $\phi : \Omega \to \mathbb{R}$ be a smooth map. Show that $\nabla^2(\phi \circ f) = |f'|^2 \nabla^2(\phi) \circ f.$

4. Obtain the following polar coordinate forms:

$$\frac{\partial}{\partial z} = \frac{1}{2} e^{-\imath\theta} \left(\frac{\partial}{\partial r} - \frac{\imath}{r} \frac{\partial}{\partial \theta} \right); \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} e^{\imath\theta} \left(\frac{\partial}{\partial r} + \frac{\imath}{r} \frac{\partial}{\partial \theta} \right).$$

3.5 Geometric Interpretation of Holomorphy

The Cauchy-Riemann equations can be viewed as saying that the tangent vectors (u_x, u_y) to the curve $u = k_1$ and the tangent vector (v_x, v_y) to the curve $v = k_2$ are orthogonal to each other. Thus indeed, it follows that the level curves of real and imaginary part of a holomorphic function are orthogonal to each other. This observation suggests another approach to holomorphicity.

Let us now try to understand the complex differentiability in a geometric way. The increment theorem or the linear approximation tells us that the behavior of a complex differentiable function in a small neighborhood of a point z_0 is approximately represented by its derivative $f'(z_0)$. Here it is essential to think of $f'(z_0)$ as the linear map $\mathbb{C} \to \mathbb{C}$ given by the multiplication $h \mapsto f'(z_0)h$. Geometrically, we know that the multiplication by a complex number has a rotational part and a scaling part. Thus a complex differentiable function must be effecting a rotation as well as a scaling at each point. Let us study this thoroughly.

First let us consider the simplest maps, viz., $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ which are real linear. Let the map be given by $T(x, y) = (ax + by, cx + dy)$. We have seen in the proof of (3.3.1), the condition under which f is complex linear, viz., $a = d$ and $b = -c$. This turns out to be the condition under which T is holomorphic as well. Moreover, we also know that a complex linear map $f : \mathbb{C} \longrightarrow \mathbb{C}$ can be described as a rotation through an angle θ followed by a (real) scaling. Such linear maps are also called 'similarities'. This leads us to consider the notion of angle preserving maps.

Recall that the standard inner product on $\mathbb{R}^n$ is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_j x_j y_j.$$

For $\mathbf{x} \neq 0, \mathbf{y} \neq 0$, the *angle* θ between them is given by

$$\cos\,\theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

For $n = 2$, in terms of complex numbers the inner product can also be expressed as $\langle z_1, z_2 \rangle = \Re(z_1 \bar{z}_2)$. Observe that our definition of the angle is insensitive to the 'direction' in which it is measured, since $\cos\,(-\theta) = \cos\,\theta$.

Definition 3.5.1 We say a linear map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *angle preserving* if it is injective and

$$\frac{\langle T(\mathbf{x}), T(\mathbf{y}) \rangle}{|T(\mathbf{x})|\ |T(\mathbf{y})|} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{|\mathbf{x}|\ |\mathbf{y}|}$$

for all non zero vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$.

Lemma 3.5.1 Let $T : \mathbb{C} \longrightarrow \mathbb{C}$ be a $\mathbb{R}$-linear map. Then the following three conditions are equivalent.
(i) T is (injective and) angle preserving.
(ii) There exists $\lambda \in \mathbb{C}^\star := \mathbb{C}\setminus\{0\}$ such that $T(z) = \lambda z,\ \forall\, z \in \mathbb{C}$ or $T(z) = \lambda\bar{z},\ \forall\, z \in \mathbb{C}$.
(iii) There exists $s > 0$ such that $\langle T(w), T(z)\rangle = s\langle w, z\rangle,\ \forall\ z, w \in \mathbb{C}$.
(iv) There exists $s > 0$ such that $|T(z)|^2 = s|z|^2,\ \forall\ z \in \mathbb{C}$.

Proof: (i) $\Longrightarrow$ (ii) Take $\lambda = T(1)$. Since T is injective, $\lambda \in \mathbb{C}^\star$. Consider the linear map $T_1(z) = \lambda^{-1}T(z)$. It is enough to show that either T_1 is the identity map or the complex conjugation. Observe that $T_1(1) = 1$. Hence it suffices to show that $T_1(\imath) = \pm\imath$. Also observe that T_1 is angle-preserving. Let $T_1(\imath) = \mu$. Then

$$0 = \langle 1, \imath\rangle = \langle T_1(1), T_1(\imath)\rangle = \langle 1, \mu\rangle.$$

Therefore $\mu = \pm r\imath$, for some $r \in \mathbb{R}^\star$. Next consider the angle between the vectors $\overrightarrow{01}$ and $\overrightarrow{1\imath}$. We have

$$\frac{\langle 1, \imath - 1\rangle}{\sqrt{2}} = \frac{\langle 1, r\imath - 1\rangle}{\sqrt{r^2+1}}$$

and this implies that $\sqrt{2} = \sqrt{r^2+1}$ and hence $r = \pm 1$, as required.
(ii) $\Longrightarrow$ (iii) Take $s = |\lambda|^2$.
(iii) $\Longrightarrow$ (iv) Take $z = w$.
(iv) $\Longrightarrow$ (i) $|T(z)|^2 = s|z|^2, s \neq 0$ and hence, if $z \neq 0$ then $T(z) \neq 0$. Since T is $\mathbb{R}$-linear, this implies T is injective. Indeed we shall first claim that (iv) $\Longrightarrow$ (iii). We begin with

$$\langle T(z+w), T(z+w)\rangle = |T(z+w)|^2 = s|z+w|^2 = s\langle z+w, z+w\rangle$$

which in turn yields

$$\langle T(z), T(z)\rangle + 2\langle T(z), T(w)\rangle + \langle T(w), T(w)\rangle = s\langle z, z\rangle + 2s\langle z, w\rangle + s\langle w, w\rangle.$$

We can now cut down the first and the last term from both sides to get condition (iii). Finally,

$$\frac{\langle T(z), T(w)\rangle}{|T(z)||T(w)|} = \frac{s\langle z, w\rangle}{\sqrt{s\langle z, z\rangle s\langle w, w\rangle}} = \frac{\langle z, w\rangle}{|z||w|}.$$

This proves (iv) $\Longrightarrow$ (i). ♠

Definition 3.5.2 Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be an injective linear map. We say T *preserves orientation* if the determinant of a matrix representing T is positive. It is straight forward to check that this property does not depend upon the choice of the matrix representing T. Also the motivation for this definition is very straight forward: *our experience tells us that maps such as rotations have positive determinant and not only preserve the angle but preserve the sense in which the angle is measured, whereas maps such as complex conjugation, reflection through any line etc. even though preserve the angle, change the sense in which the angle is measured.*

Example 3.5.1 The maps $x \mapsto -x$, $(x, y) \mapsto (y, x)$, $(x, y, z) \mapsto (-x, -y, -z)$ are all orientation reversing. The map $(x_1, \ldots, x_n) \mapsto (-x_1, \ldots, -x_n)$ is orientation preserving iff n is odd. The maps $(x, y) \mapsto (ax + by, cx + dy)$ preserves orientation iff $ad - bc > 0$.

Remark 3.5.1 The problem that we are presently interested in is how to extend this concept to maps that are not necessarily linear. We must first of all define angle between a pair of intersecting 'curves'.

Definition 3.5.3 By a 'smooth curve' in $\mathbb{R}^n$ we mean a continuously differentiable map $\gamma : I \longrightarrow \mathbb{R}^n$ such that $\gamma'(t) \neq 0,\ t \in I$ where I denotes some interval. For any two smooth curves passing through a given point p, we define the angle between them at p to be the angle between the tangents to the two curves at p.

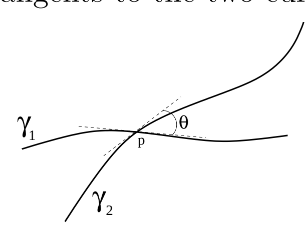


Fig. 11

Finally, let U be an open subset of $\mathbb{R}^m, A \subset U$ and $f : U \longrightarrow \mathbb{R}^n$ be a continuously differentiable function. We say $f|_A$ is angle preserving at $p \in A$ if for every pair γ_1, γ_2 of smooth curves in A passing through p, we have the angle between them at p is equal to the angle between the image curves $f \circ \gamma_1$ and $f \circ \gamma_2$.

The following lemma is immediate from the chain-rule.

Lemma 3.5.2 A differentiable map f defined on an open subset U of $\mathbb{R}^m$ is angle preserving at $p \in U$ iff Df_p is angle preserving as a linear map. It is orientation preserving at p iff Df_p is so.

Example 3.5.2 Any linear map given by a multiplication by a non zero complex number is angle as well as orientation preserving. The complex conjugate is a $\mathbb{R}$-linear isomorphism which does not preserve orientation. We also say, in this case that the map *reverses orientation.*

Theorem 3.5.1 *Let $f : U \longrightarrow \mathbb{C}$ be any map on an open subset U of $\mathbb{C}$. Then the following are equivalent:*
(i) *f is complex differentiable on U and f' does not vanish on U.*
(ii) *f is real differentiable, $Df(z) \neq 0$, for any $z \in U$, and f preserves angle and orientation at each point of U.*

Proof: (i) $\implies$ (ii) Since f is $\mathbb{C}$-differentiable and $f'(z) \neq 0$, it follows that f is real differentiable with $(Df)_z = f'(z) \neq 0$. Moreover, $f'(z)$ is given by multiplication by a complex number and hence by the above lemma, f is angle preserving. That it preserves orientation also follows from the fact that multiplication by a non zero complex number does so.
(ii) $\implies$ (i) Again by the lemmas 3.5.1,3.5.2, $(Df)_z(w) = \lambda w$ for all w or $(Df)_z(w) = \lambda \bar{w}$ for all w. The latter possibility is ruled out because of the orientation preserving property. Thus $(Df)_z$ is complex linear, and from theorem 3.3.1, we know that f is $\mathbb{C}$-differentiable at z. Since this is the case at all points $z \in U$ we are done. ♠

Definition 3.5.4 A smooth map $f : U \to \mathbb{R}^n$ where U is an open subset of $\mathbb{R}^n$ is called *conformal* at $p \in U$ if it preserves angle and orientation at p. (Note that this assumes that Df_p is an isomorphism.)

Remark 3.5.2 (i) Thus any holomorphic function is conformal at all points where its derivative is not zero.
(ii) The relation between conformality and holomorphicity is much closer than what we have seen above. First suppose that $U = B_r(z_0)$ is an open disc in $\mathbb{R}^2$. Let $f : U \longrightarrow \mathbb{R}^2$ be a continuously real differentiable, angle preserving function. Define $\phi : t \mapsto (1 - t)z_0 + tz$ and take $g(t) = \det(Df_{\phi(t)})$. Then, since f is continuously differentiable, and taking determinant is a continuous operation, it follows that g is a continuous map. Since f is angle preserving, $\det(Df)$ does not vanish at any point. Therefore, g does not assume the value 0. By intermediate value theorem, it follows that $g(0)(= Df_{z_0})$ and $g(1)(= Df_z)$ have the same sign. Thus if f is orientation preserving at z_0, then it is so all over U and hence by the above theorem, it is holomorphic in U. On the other

hand, if f is orientation reversing at z_0, $\bar{f}$ is orientation reversing all over U and hence f is anti-holomorphic.

In the general case, we choose a path ϕ from z_0 to any given point z inside U and argue with $g(t) = Df_{\phi(t)}$.

(iii) The scaling factor is much easy to understand. At each point z_0 where the derivative of f is not zero, we see that

$$\lim_{z\to z_0} \frac{|f(z) - f(z_0)|}{|z - z_0|} = \left|\lim_{z\to z_0} \frac{f(z) - f(z_0)}{z - z_0}\right| = |f'(z_0)|$$

represents the factor by which the distance from z_0 to a nearby point gets expanded.

Exercise 3.5

1. Determine the points at which the following maps preserve angles:
   a) $z \mapsto z^2$; b) $z \mapsto \sin z$; c) $z \mapsto \sqrt{z}$, $\Re(z) > 0$.

2. Consider the polar co-ordinate mapping

   $$g(r,\theta) = (r\cos\theta, r\sin\theta),\ 0 < r < \infty,\ -\infty < \theta < \infty.$$

   Determine all points where g is conformal.

3. Determine the rotational and scaling factors of $z \mapsto z^2$ at the points
   (i) $\sqrt{2}$; (ii) $\imath$; (iii) $1 + \imath$.

4. Let $f : U \to \mathbb{R}^2$ be smooth mapping of a domain in $\mathbb{R}^2$ with $det\, Df_p \neq 0$. Suppose that at each point it is angle preserving and the angle of rotation is a constant throughout the domain. Show that f is of the form $z \mapsto az + b$; or $z \mapsto a\bar{z} + b$. Can you conclude the same thing if the scaling factor is a constant throughout in U?

5. Suppose $w = f(z)$ is a one-to-one, conformal mapping of a domain D_1 in the xy-plane onto a domain D_2 in the uv-plane. Let $\phi(u, v)$ be a real valued function with continuous partial derivatives on D_2 and let ψ be the composite function $\phi \circ f$ on D_1. Show that
   (i) The level curves of ψ are mapped onto the level curves of ϕ by f, i.e., ϕ is a constant on C_2 if and only if ψ is a constant on C_1.
   (ii) The normal derivative of ϕ along C_2 vanishes iff the normal derivative of ψ along C_1 vanishes.

(iii) In general it is not true that the normal derivative of ψ along C_2 is a constant if normal derivative of ϕ along C_1 is a constant. [Hint: Consider $f(z) = z^2$.]

3.6 Mapping Properties of Elementary Functions

There are many situations in which we have to know specific behavior of a given holomorphic function on a restricted domain. Such a study cannot be exhaustive by no means. In this section, we present this aspect through a few well-chosen examples. A common feature of all this study is conformality. Thus, we first of all discard the set of all points where the derivative of the function vanishes. Next, we somehow ensure that the function is one-to-one by restricting the domain if necessary. By lemma 2.4.1, it follows that if we have a continuous inverse of the function, then it is automatically holomorphic. A holomorphic function f with a holomorphic inverse is called a *biholomorphic mapping;* the domain of f and its image are then said to be *biholomorphically equivalent.*

Thus, in particular, we are going to witness biholomorphic equivalence between various domains in $\mathbb{C}$.

Example 3.6.1 Linear Functions Let us begin with the simplest examples, viz., linear functions $z \mapsto az + b$. If $a = 0$ then we get a constant function which is not of much interest. If $|a| = 1$, then we have seen that this is a rigid motion which is a composite of a rotation around 0 followed by a translation. These mappings have been used so often that we do not even notice them any longer. Every time we choose the origin, we are implicitly making a translation and the choice of the x-axis and y-axis means that we perform a rotation.

In the general case, there is a scaling factor $|a|$. Nevertheless the mapping is conformal everywhere, all straight lines are mapped to straight lines and circles are mapped to circles. Thus rotating, translating and enlarging etc. of a domain can be achieved by linear mappings. Later, we shall see that there is no biholomorphic mappings of $\mathbb{C}$ other than the linear ones.

Example 3.6.2 The Square Function and the Square-root Function: Now, consider the simplest polynomial mapping of higher degree: $w = z^2$. The derivative of this map is non zero everywhere except at $z = 0$. Thus the mapping is conformal in $\mathbb{C}^\star$.

Clearly, it is not injective on $\mathbb{C}^\star$ and so let us first take a domain on which it is so. On any sector whose span is of angle $< \pi$, this function is injective. In particular, the open first quadrant is mapped biholomorphically onto the upper half-plane (along with the boundary). Also if we restrict the function to the upper-half plane

$$\boldsymbol{H} := \{z \in \mathbb{C} : \Im z > 0\}$$

then it is injective. The image of $\boldsymbol{H}$ under this map is $\mathbb{C} \setminus \mathbb{R}^+$ obtained by deleting all the non negative real numbers from the plane. Verify likewise that if L is any line passing through 0, then $w = z^2$ restricted to any one side of the line is one-one. Putting $w = u + \imath v$ and $z = x + \imath y$ we see that

$$u = x^2 - y^2; \quad v = 2xy$$

Thus, each lap of the hyperbola $x^2 - y^2 = c$ is mapped onto the vertical line $u = c$. Similarly, each lap of the hyperbola $xy = k$ is mapped into the horizontal line $v = 2k$. From this, it is easy to determine the image of rectangles with sides parallel to x and y axis and contained in any quadrant under the mapping $z \mapsto \sqrt{z}$.

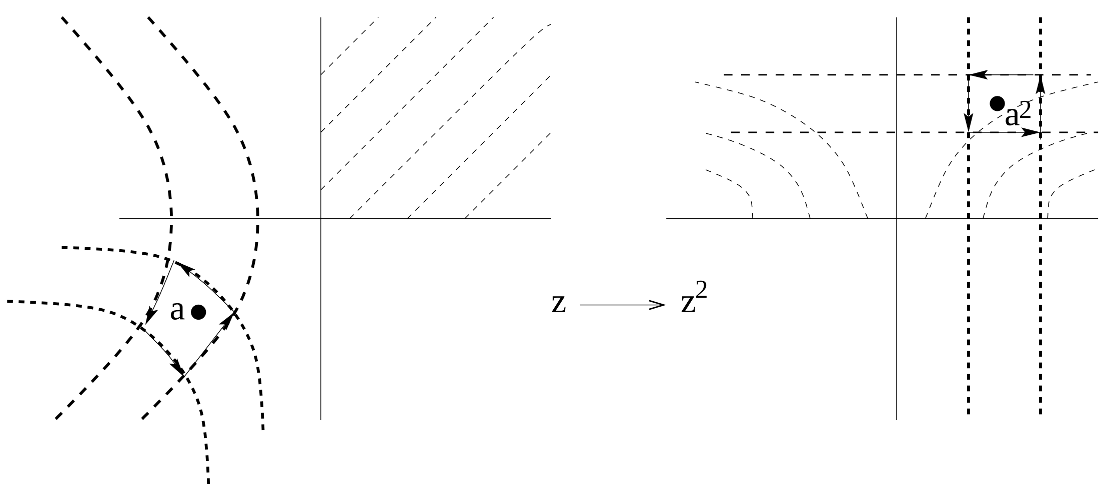


Fig. 12

Example 3.6.3 The Exponential and the Logarithm : Consider the map exp : $\mathbb{C} \longrightarrow \mathbb{C}^\star$ given by $z \mapsto e^z$. Since $(e^z)' = e^z \neq 0$, this is a conformal mapping everywhere. Observe that the image of the real axis under this map is the positive real axis, whereas the image of the imaginary axis is the unit circle. The point 0 is mapped to the point 1. The two axes intersect at 0 orthogonally. So do their images under exp,at the point

1. Similarly, all horizontal lines are mapped onto radial half-rays originating from 0. All vertical lines wind around onto circles with center at the origin infinitely many often. The image of the line $y = x$ is a spiral. It is easily checked that on the open infinite strip $0 < \Im z < \pi$ the function e^z in univalent (i.e., injective), the image being the upper-half plane $\boldsymbol{H}$. Also check that the portion of the imaginary axis between 0 and $\imath\pi$ is mapped onto the upper part of the unit circle. The parts of the strip with $\Re(z) > 0$ (resp. < 0) are mapped outside (inside) of the unit semi-disc.

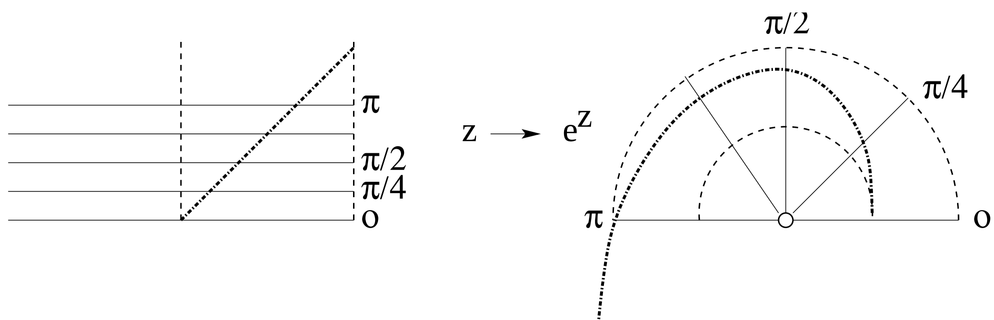


Fig. 13

Example 3.6.4 The Sine Function : This map is conformal at points other than $z = \pi(2m+1)/2$.

What is the image? Put $T = e^{\imath z}$. Given w finding z such that $\sin z = w$ is the same as finding T such that

$$\frac{T - T^{-1}}{2\imath} = w.$$

This is a quadratic in T which has two solutions none of which is equal to 0. Therefore, $\sin z$ is onto.

Observe that $\sin(z+\pi) = -\sin z = \sin(-z)$. Therefore to get a one-one function, the function has to be restricted to $U = \{(x,y) : -\pi/2 \leq x \leq \pi/2\}$ or to $V = \{(x,y) : -\pi < x < \pi; y > 0\}$. Then of course, it is a one-one mapping.

Writing $z = x + \imath y$ and $\sin z = u + \imath v$, we have

$$u = \sin x \cosh y; v = \cos x \sinh y.$$

Observe that the two boundary lines $x = \pm\pi/2$ are mapping into the horizontal line $v = 0$. It follows that the vertical lines $x = c \neq 0$ are mapped to the hyperbolas

$$\frac{u^2}{\sin^2 c} - \frac{v^2}{\cos^2 c} = 1 \tag{3.25}$$

and the horizontal lines $y = k$ are mapped onto ellipses

$$\frac{u^2}{\cosh^2 k} + \frac{v^2}{\sinh^2 k} = 1. \tag{3.26}$$

These two families are obviously orthogonal to each other. All these hyperbolas and ellipses have the same foci: $\pm\sqrt{\cos^2 c + \sin^2 c} = \pm 1; \pm\sqrt{\cosh^2 k - \sinh^2 k} = \pm 1$. Thus, it is easily seen that a rectangle of the form $[-\pi/2, \pi/2] \times [k_1, k_2], k_1 > 0$ is mapped onto the region between two ellipses, in a one-one fashion except that both vertical sides are mapped on to the same cross-cut along the negative imaginary axis.

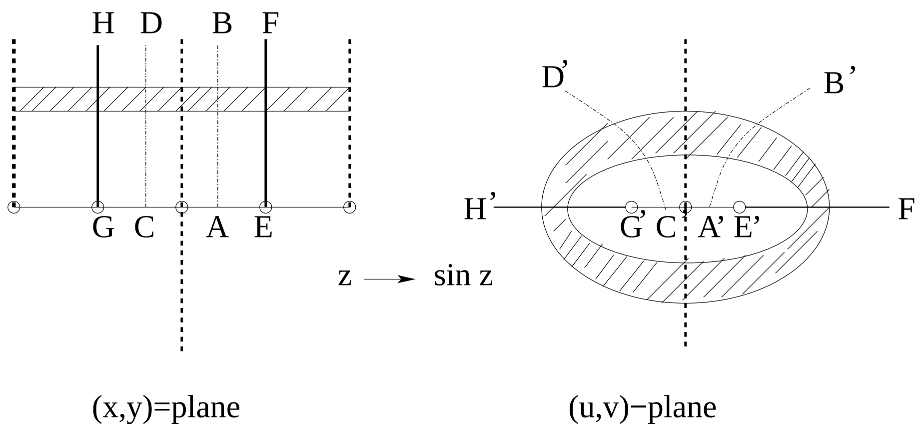


Fig. 14

The cosine function does not offer any variety, since it is got by mere translation of $\sin z$, i.e., $\cos z = \sin(z + \pi/2)$.

Our next example is going to be a special case of rational functions. It is a very important class of holomorphic functions and so we shall study this class in a separate section.

Exercise 3.6

1. Find the image of the square $\{(x, y) \ : \ 0 < a \leq x \leq b, 0 < c \leq y \leq d\}$ under the mapping $z \mapsto \sqrt{z}$.

2. Under the mapping $z \mapsto e^z$, determine the image of
   (i) the rectangle $\{(x, y) \ : \ a \leq x \leq b, c \leq y \leq d\}$;
   (ii) the lines $x = my$.
   (iii) the semi-infinite strip $\{(x, y) \ : \ x \geq 0, 0 \leq y \leq \pi\}$;
   (iv) the sector $\{(r, \theta) \ : \ \theta_1 \leq \theta \leq \theta_2\}$.

3. Show that the transformation $w = \sin z$ maps the line $x = \phi, \ \pi/2 < \phi < \pi$ in a bijective manner onto the lap of a hyperbola lying on the right half plane.

4. Determine the image of the infinite semi-strip $\{(x, y)\ :\ -\pi/2 < x < \pi/2, y > 0\}$ under the transformation $w = \sin z$. Verify whether it is one-one.

5. Determine the image of the semi-infinite strip $\{(x, y)\ :\ x > 0, 0 < y < \pi/2\}$ under the mapping $w = \cosh z$. [Hint Use the formula: $\sin(\imath z + \pi/2) = \cosh z$.]

3.7 Fractional Linear Transformations

Definition 3.7.1 A fractional linear transformation (flt) (also called Möbius transformation) is a *non constant* rational function in which both numerator and denominator are at most of degree one in z. Thus they are given by the formula

$$z \mapsto \frac{az + b}{cz + d} \tag{3.27}$$

Remark 3.7.1

1. Of course at least c or d has to be non zero in order to make any sense out of this formula. Also for $c = 0$ this define a linear map $z \mapsto \frac{a}{d}z + \frac{b}{d}$, and whatever we are going to say about fractional linear transformation is easily verified in that case. So, throughout, we shall assume that $c \neq 0$ whenever we are saying something about a particular fractional linear transformation, though, the collection of all fractional linear transformation would of course contain all linear maps as well. (Observe that the word 'linear' is used here in the larger sense of *affine linear*.) In any case, we shall never consider constant maps in the following discussion. We could have included the constant maps in the definition of fractional linear transformations for a formalistic reason, but they do not possess any of the geometric properties that we are interested in.

2. The formula (3.27) makes sense for all points $z \neq -d/c$. Also it defines a holomorphic function with its derivative

$$\frac{ad - bc}{(cz + d)^2}.$$

   This suggests that we could associate a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with the above map, so that the numerator of the derivative is the determinant of A. Then to say that (3.27) defines a non constant map is the same as saying that det $A \neq 0$. We let $GL(2, \mathbb{C})$ denote the set of all non singular 2×2 matrices over the complex numbers and observe that this forms a group under matrix multiplication. Now the most interesting thing is that if we assign to each $A \in GL(2, \mathbb{C})$ the fractional linear transformation h_A defined as in (3.27), then the assignment $A \mapsto h_A$ is a homomorphism:

$$h_{A \circ B} = h_A \circ h_B$$

for all $A, B \in GL(2, \mathbb{C})$. (Verify this). In particular, it follows that $h_A \circ h_{A^{-1}} = h_{A \circ A^{-1}} = h_I = Id$.

3. Thus, all (non constant) fractional linear maps are invertible. In particular, they are injective. Observe that it is not stated that a fractional linear transformation is surjective onto $\mathbb{C}$. As we have observed before, a fractional linear transformation may not be even defined at a point ($z = -d/c$). Since the inverse of a fractional linear transformation makes sense at all points of $\mathbb{C}$ except perhaps one, it follows that exactly one point of the complex plane is missing from the image of any fractional linear transformation. Moreover, from theorem 3.5.1, it follows that each fractional linear transformation is conformal. Thus

$$h_A : \mathbb{C} \setminus \{-dc^{-1}\} \longrightarrow \mathbb{C} \setminus \{ac^{-1}\}$$

is a biholomorphic mapping. The 'missing points' from the domain as well as from the codomain of a fractional linear transformation can be taken care of in a nice way. We need a definition.

Definition 3.7.2 By the *extended complex plane* we mean $\mathbb{C} \cup \{\infty\}$ and denote it by $\widehat{\mathbb{C}}$.

Remark 3.7.2

1. For more about the extended complex plane which is also called the Riemann sphere, see the next section.

2. We can now continue with the discussion of fractional linear transformation in the extended complex plane as well. First of all it now makes sense to assign the value ∞ to $h_a(z)$ for $z = -d/c$. This is justified by the fact

$$\lim_{z \to -d/c} \frac{az+b}{cz+d} = \infty.$$

Likewise it makes sense to talk about $h_A(\infty)$ viz.,

$$h_A(\infty) = \lim_{z\to\infty} h_A(z) = \lim_{z\to\infty} \frac{az+b}{cz+d} = \frac{a}{c}.$$

It follows that each non constant fractional linear transformation $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a one-one and onto mapping.

3. We next observe that

$$h_A(z) = \frac{a}{c} + \left(\frac{-ad+bc}{c^2}\right)\left(\frac{1}{z+d/c}\right).$$

Thus it is clear that we can write h_A as a composite of a few very simple maps: Let T_α denote translation by α viz., $z \mapsto z + \alpha$. Similarly let μ_α denote the multiplication by α. Finally let η denote the inversion $z \mapsto z^{-1}$. Put $d/c = \lambda_1$, $(bc-ad)/c^2 = \lambda_2$ and $a/c = \lambda_3$. Then we see that

$$h_A = T_{\lambda_3} \circ \mu_{\lambda_2} \circ \eta \circ T_{\lambda_1}. \tag{3.28}$$

Since, the geometric behavior of translations rotations and scaling are easily understood, in order to understand the geometric properties of h_A, we have to study the geometric properties of the inversion map η alone. As an illustration let us prove:

Theorem 3.7.1 *The set of all circles and straight lines in the plane is preserved by any fractional linear transformation.*

Proof: (Observe that the theorem does not assert that each circle is mapped to a circle. Nor does it say that each line is mapped to a line.) From the decomposition (3.28) for a fractional linear transformation, it is clear that we need to verify this property only for the map η. Because any way the other maps involved in the composition are orthogonal transformations, translations or scaling, which map circles to circles and lines to lines.

Now recall from your high school geometry that an equation of the form

$$\alpha(x^2+y^2) + \beta x + \gamma y + \delta = 0 \tag{3.29}$$

(where $\alpha, \beta, \gamma, \delta \in \mathbb{R}$) represents a circle (or a straight line) if $\alpha \neq 0$ (respectively, if $\alpha = 0$.) If $z := x + \imath y \neq 0$ and $w := z^{-1} = u + \imath v$ then we have

$$u = \frac{x}{x^2+y^2}; \; v = \frac{-y}{x^2+y^2}.$$

$$x = \frac{u}{u^2+v^2}; \; y = \frac{-v}{u^2+v^2}.$$

Therefore, $z = x + \imath y$ satisfies (3.29) iff $w = u + \imath v$ satisfies

$$\delta(u^2+v^2) + \beta u - \gamma v + \alpha = 0. \tag{3.30}$$

This last equation represents a circle or a straight line according as $\delta \neq 0$ or $= 0$. ♠

Remark 3.7.3

1. It is also clear that when a circle is mapped onto a straight line and vice versa by a fractional linear transformation h_A, viz., circles which pass through $-d/c$ and straight lines which do not pass through $-d/c$. (For a better understanding of this, read section 3.8).

2. **Bilinearity** Another aspect of the fractional linear transformation is that we can write it by an implicit equation of the form

$$cwz + dw + az + b = 0 \tag{3.31}$$

where, $ad - bc \neq 0$. (What we have done is to put $w = -\dfrac{az+b}{cz+d}$ and simplify.) The formula (3.31) can be used to define both the transformation and its inverse: $z \mapsto w; \;\; w \mapsto z$. Observe that the above equation is a polynomial equation in two variables $z, \; w$; it is a linear polynomial in each of the variables. That is the reason why a fractional linear transformation is also called a *bilinear transformation* in literature. Note that the fractional linear transformation $z \mapsto w$ defined by (3.31) is uniquely determined by the vector (c, d, a, b) and any non zero multiple of this vector also defines the same fractional linear transformation.[3]

3. We know that a real linear map on $\mathbb{R}^2$ is completely determined by its value on any two independent vectors. In the case of fractional linear transformations, the situation is similar. For, suppose a fractional linear transformation T given by

[3]This means that the set of all non constant fractional linear transformations forms an *affine open subset of the 3-dimensional complex projective space.*

(3.27) fixes a point w. This means that w satisfies the equation $cX^2+(d-a)X-b = 0$ which is a polynomial equation of degree ≤ 2. Now assume that T fixes three distinct points. Since any polynomial of degree less than or equal to 2 with three distinct roots has to be identically zero, we get, $c = 0 = b$ and $a = d$. This is the same as saying that T is the identity map. Thus we have proved:

Theorem 3.7.2 *Every fractional linear transformation which fixes three distinct points of $\widehat{\mathbb{C}}$ is necessarily the identity map.*

Remark 3.7.4

1. Observe that, in general, the fixed points of a fractional linear transformation satisfy a quadratic and hence, there are two of them.

2. We can now conclude that if two fractional linear transformations agree on any three distinct points then they must be the same. For, if $T_1(z_j) = T_2(z_j), j = 1, 2, 3,$ then it follows that $T_1^{-1} \circ T_2(z_j) = z_j, i = 1, 2, 3$. Therefore, $T_1^{-1} \circ T_2 = Id$.

3. Finally, we must see whether we can have a fractional linear transformation which maps given three distinct points at our will to three other points. Writing

$$w = \frac{az + b}{cz + d}$$

   and substituting $w = w_j, z = z_j, j = 1, 2, 3$, we obtain three linear equations in four unknowns (a, b, c, d). Therefore, we certainly have non zero solutions. But we want a solution in which either a or c is not zero. Check that this condition is easily satisfied for otherwise, all the three points w_j will be the same. We shall however, include a more formal proof of this very important result below:

Theorem 3.7.3 *Given two sets $\{z_j\}$ and $\{w_j\}$ of three distinct elements each in $\widehat{\mathbb{C}}$, there is a unique fractional linear transformation f such that $f(z_j) = w_j, j = 1, 2, 3$.*

Proof: We first consider a special case, when $z_1 = -1, z_2 = 0$ and $z_3 = 1$ and $w_j \in \mathbb{C}, j = 1, 2, 3$. Plugging these values in (3.27) we get,

$$\text{(i)} \quad \frac{-a+b}{-c+d} = w_1; \quad \text{(ii)} \quad \frac{b}{d} = w_2; \quad \text{(iii)} \quad \frac{a+b}{c+d} = w_3.$$

If all w_j are in $\mathbb{C}$ we can simply take $d = 1$ and solve the three linear equations for a, b, c. If $w_1 = \infty$ then we choose $c = d = 1$ and solve (ii) and (iii) for a, b. If $w_2 = \infty$, we take

$d = 0, c = 1$ and solve (i) and (iii) for a, b. Finally, if $w_3 = \infty$ we take $c = -d = 1$ and solve (i) and (ii) for a, b.

In the general case, we let T, S be fractional linear transformations such that $T(-1) = w_1, T(0) = w_2, T(1) = w_3$ and $S(-1) = z_1, S(0) = z_2, S(1) = z_3$. Then it follows that $T \circ S^{-1}(z_j) = w_j, j = 1, 2, 3$. Since, inverse of a fractional linear transformation is a fractional linear transformation and composite of two fractional linear transformation is a fractional linear transformation, $T \circ S^{-1}$ is a fractional linear transformation. ♠

Remark 3.7.5 Symmetric form of FLT Another interesting way of putting the fractional linear transformation $S \circ T^{-1}$ is the following: Write $S \circ T^{-1}(w) = z$. This is the same as $T^{-1}(w) = S^{-1}(z)$. In the form (3.27), this reads as

$$\frac{az + b}{cz + d} = \frac{a'w + b'}{c'w + d'} \tag{3.32}$$

This can now be rewritten in the form

$$\alpha \cdot \frac{z - \beta}{z - \gamma} = \alpha' \cdot \frac{w - \beta'}{w - \gamma'} \tag{3.33}$$

The idea is that z and w are expressed in a symmetrical fashion in this formula, so that it can be thought of as a mapping $z \mapsto w$ or its inverse $w \mapsto z$. Of course, here $\alpha, \alpha' \neq 0$. As an application, we can now explicitly determine the fractional linear transformation which maps z_j to $w_j, j = 1, 2, 3$ respectively as follows:

By choosing $\beta = z_1$, we see that lhs of (3.33) vanishes at $z = z_1$. This suggests that we should choose $\beta' = w_1$. For the same reason, we choose $\gamma = z_2, \gamma' = w_2$. Finally, in order to satisfy the condition that when $z = z_3$, we have $w = w_3$, we merely plug these choices in (3.33) to obtain

$$\alpha \cdot \frac{z_3 - z_1}{z_3 - z_2} = \alpha' \cdot \frac{w_3 - w_1}{w_3 - w_2}.$$

This determines the value of α/α'. For the sake of symmetry, we choose $\alpha = \dfrac{z_3 - z_2}{z_3 - z_1},\ \alpha' = \dfrac{w_3 - w_2}{w_3 - w_1}$. Therefore we get the required fractional linear transformation :

$$\boxed{\frac{(z - z_1)(z_3 - z_2)}{(z - z_2)(z_3 - z_1)} = \frac{(w - w_1)(w_3 - w_2)}{(w - w_2)(w_3 - w_1)}} \tag{3.34}$$

Example 3.7.1 To illustrate the algorithmic nature of (3.34), let us consider the problem of determining the fractional linear transformation that maps $0 \mapsto 1,\ 1 \mapsto \imath$ and $-1 \mapsto -\imath$. We simply write

$$\frac{(z-0)(-1-1)}{(z-1)(-1-0)} = \frac{(w-1)(-\imath-\imath)}{(w-\imath)(-\imath-1)}.$$

Upon simplification this turns out to be

$$w = \frac{\imath - z}{\imath + z}.$$

The symmetric format (3.34) of a fractional linear transformation also relates it to another classical geometric notion.

Definition 3.7.3 Given four distinct points z_1, z_2, z_3, z_4 of $\widehat{\mathbb{C}}$, we define their **cross ratio** to be

$$[z_1, z_2, z_3, z_4] = \left(\frac{z_1 - z_3}{z_1 - z_4}\right) / \left(\frac{z_2 - z_3}{z_2 - z_4}\right) = \left(\frac{z_1 - z_3}{z_2 - z_3}\right)\left(\frac{z_2 - z_4}{z_1 - z_4}\right).$$

(See also Exercise 1.9.26.) Here if one of the points is ∞ then the meaning assigned to $[z_1, z_2, z_3, z_4]$ is to replace ∞ by a complex number z and take the limit as $z \to \infty$. For example,

$$[\infty, z_2, z_3, z_4] = \lim_{z \to \infty} \left(\frac{z - z_3}{z - z_4}\right) / \left(\frac{z_2 - z_3}{z_2 - z_4}\right) = \frac{z_2 - z_4}{z_2 - z_3}.$$

Remark 3.7.6

1. Observe that the order in which you take the four numbers is important. It is an interesting exercise to find out how the cross ratios are related under permutation of the four numbers.

2. Fixing z_2, z_3, z_4 the map $z \mapsto [z, z_2, z_3, z_4]$ is a fractional linear transformation. Conversely, given z_2, z_3, z_4 the fractional linear transformation which takes

$$z_2 \mapsto \frac{z_2 - z_3}{z_2 - z_4}; \quad z_3 \mapsto 0; \quad z_4 \mapsto \infty$$

is nothing but $[z, z_2, z_3, z_4]$. Thus fractional linear transformations are nothing but cross-ratios. Of course, each cross ratio gives rise to different fractional linear transformations depending upon which one of the four slots is treated as a free variable and the other three fixed.

Theorem 3.7.4 *Let T be a fractional linear transformation. Then*

$$[T(z_1), T(z_2), T(z_3), T(z_4)] = [z_1, z_2, z_3, z_4].$$

Proof: Since T is the composite of translation, rotation, dilation and inversion, it is enough to prove this statement when T itself is one of these. The case when $T(z) = 1/z$ is the one which is non trivial. Even this is routine and hence we leave this as an exercise. ♠

Finally we shall discuss one important example of fractional linear transformation, the negative of the above fractional linear transformation.

Example 3.7.2 The Cayley[4]map $\chi : \boldsymbol{H} \longrightarrow \mathbb{D}$:

Introduce the notation

$$\boldsymbol{H} = \{z \ : \ \Im(z) > 0\}; \quad \mathbb{D} := \{z \ : \ |z| < 1\}.$$

Consider the map $\chi : z \mapsto \dfrac{z - \imath}{z + \imath}$. It is not defined only at $z = -\imath$. Since, $|z+\imath|^2 - |z-\imath|^2 = |z|^2 + |\imath|^2 + 2\,\Re(z\bar{\imath}) - |z|^2 - |\imath|^2 - 2\Re(-z\bar{\imath}) = 4\Im(z)$, it follows that $\Im z > 0$ iff $|\chi(z)| < 1$. Moreover, the inverse map of χ is given by

$$w \mapsto \frac{1 + w}{1 - w} \cdot \imath.$$

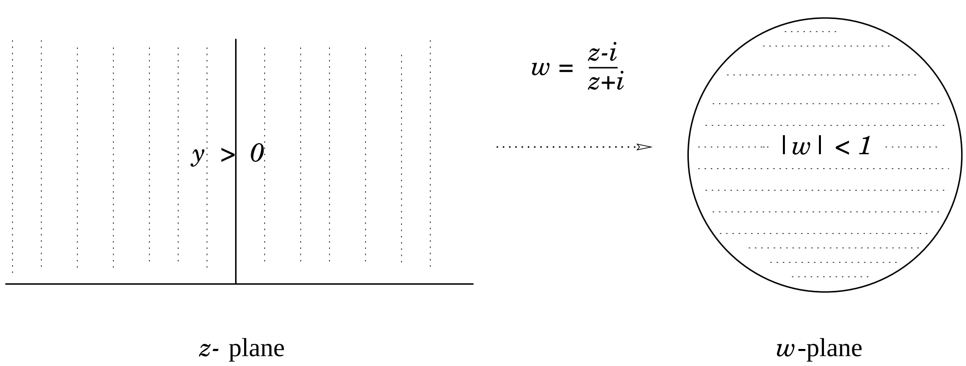


Fig. 15

Therefore χ is a biholomorphic mapping of the upper-half plane onto the open unit disc. Observe that the image of the points $-1, 0, 1$ are respectively, $\imath, -1, -\imath,$. In fact

[4]Arthur Cayley(1821-1895) an English mathematician was one of the proponents of the theory of algebraic invariants.

it is also clear that the entire real axis is mapped bijectively onto the unit circle minus the point 1. The upper part of the imaginary axis is mapped onto the interval $(-1, 1)$. It will be an interesting exercise to see for yourself what the image of any particular line or circle under χ is. Of course, we already know that each of them is either a half-line or a portion of a circle.

Remark 3.7.7 One may wonder how one might have arrived at such a remarkable mapping. Here is a probable explanation: We know that fractional linear transformations are anyway injective mappings and hence, there is a good chance of them defining biholomorphic mappings of two different domains in $\mathbb{C}$, if there are any. Next, if a fractional linear transformation say f defines a biholomorphic mapping of $\boldsymbol{H}$ onto $\mathbb{D}$, then it would also map the real axis into the unit circle. Now, if we trace the real axis in the positive direction, then $\boldsymbol{H}$ lies to our left, and, if we trace the unit circle in the counter clockwise sense, then $\mathbb{D}$ lies to our left. Since, any fractional linear transformation preserves orientation, we should choose our f so that it maps the real axis traced in the positive sense into unit circle traced in the counter clockwise sense. Given these considerations what could be better than choosing f to map $-1, 0, 1$ respectively onto $\imath, -1, -\imath$? You may choose any three distinct real numbers $r_1 < r_2 < r_3$ and map them onto $\imath, -1, -\imath$, respectively to get other fractional linear transformations and check that they too define biholomorphic mappings of $\boldsymbol{H}$ onto $\mathbb{D}$. The following theorem puts any further speculation to rest.

Theorem 3.7.5 *Let T be a fractional linear transformation mapping the open unit disc onto itself. Then T is of the form*

$$T(z) = c\frac{z-a}{1-\bar{a}z}$$

for some a, c such that $|a| < 1$ and $|c| = 1$.

Proof: First we prove that if T is of the form as above then $T(\mathbb{D}) = \mathbb{D}$. Since inverse of T is also of the above form it suffices to show $T(\mathbb{D}) \subset \mathbb{D}$. Since $|c| = 1$, it is enough to show that $|z - a| \leq |1 - \bar{a}z|$ for $|z| < 1$. Using cosine rule, this is the same as proving $|z|^2 + |a|^2 \leq 1 + |az|^2$ which in turn is the same as proving $|a|^2(1 - |z|^2) \leq 1 - |z|^2$ for $|z| < 1$. This last inequality follows since $|a| < 1$.

Now given a fractional linear transformation T which maps $\mathbb{D}$ onto itself, take $a = T^{-1}(0)$ and put

$$S(z) = \frac{z-a}{1-\bar{a}z}.$$

so that S maps a to 0 and is a fractional linear transformation which maps $\mathbb{D}$ onto itself. Therefore if $R = T \circ S^{-1}$, then R is a fractional linear transformation which maps $\mathbb{D}$ to $\mathbb{D}$ and 0 to 0. The second condition implies that R is of the form

$$R(z) = \frac{\beta z}{\gamma z + \delta}.$$

Since the unit circle has to be mapped onto the unit circle by R, it follows that $|\beta| = |\gamma z + \delta|$ for all $|z| = 1$. If $\gamma \neq 0$, then $|z + \delta/\gamma| = |\beta/\gamma|$ for every $z \in \mathbb{S}^1$, i.e., $\mathbb{S}^1$ is a circle with center $-\delta/\gamma$. This implies $\delta = 0$ and hence R is constant map which is absurd. Therefore, we conclude that $\gamma = 0$ and $|\beta| = |\gamma|$. Therefore R is a rotation. Say $R(z) = cz$ with $|c| = 1$. Therefore, $T = R \circ S$, and we get the required form for T. ♠

Exercise 3.7

1. Find the image of the first quadrant Q_1 under the Cayley map.

2. Given $|z_0| < R$, show that
$$z \mapsto \frac{R(z - z_0)}{R^2 - \bar{z}_0 z}$$
maps the disc $|z| < R$ bijectively onto $\mathbb{D}$ sending z_0 to 0. (Pay attention to the mapping in the special case $R = 1$, which gives a very useful map.)

3. Given $z_0 \in \boldsymbol{H}$ show that the transformation:
$$z \mapsto \frac{z - z_0}{z - \bar{z}_0}$$
maps $\boldsymbol{H}$ univalently onto the unit disc $\mathbb{D}$ and sends z_0 to 0.

4. Show that the mapping
$$z \mapsto \frac{1 + z}{1 - z}$$
defines an holomorphic equivalence of the unit disc $\mathbb{D}$ and the domain
$$A = \{z \ : \ \Re z > 0\}.$$

5. Determine how many distinct values of the cross ratio $[z_1, z_2, z_3, z_4]$ you obtain by permuting the four points z_1, z_2, z_3, z_4. (See Miscellaneous Exercise 1.9.26 or definition 3.7.3.)

6. Show that given any two points z_1, z_2 in the upper-half plane there exits a fractional linear transformation which maps z_1 to z_2 and maps the upper-half plane to itself. Have you seen similar result for the unit disc?

7. Show that any fractional linear transformation which is of the form

$$z \mapsto \frac{az+b}{cz+d}, \quad a,b,c,d \in \mathbb{R}, \quad ad-bc=1 \tag{3.35}$$

maps the upper-half plane onto itself and conversely any fractional linear transformation which maps the upper-half plane onto itself is of this form.

8. * Consider the following subspace X of the upper-half plane given by $\tau \in \boldsymbol{H}$ such that
(i) $|\tau| \geq 1$;
(ii) $-1/2 < \Re(\tau) \leq 1/2$;
(iii) If $|\tau| = 1$ then $\Re(\tau) \geq 0$.
Show that given any $\tau \in \boldsymbol{H}$, there is a unique $\tau' \in X$ and a fractional linear transformation A as in (3.35) such that $A\tau' = \tau$. (See theorem 10.7.2 and the accompanying figure.)

3.8 The Riemann Sphere

Consider the 2-dimensional unit sphere in $\mathbb{R}^3$:

$$\mathbb{S}^2 := \{(x_1, x_2, x_3) \ : \ x_1^2 + x_2^2 + x_3^2 = 1\}.$$

Let us denote the point $(0,0,1)$ by N and agree to call it the *north pole.* (Likewise, one calls the point $S = (0,0,-1)$ the *south pole.*)

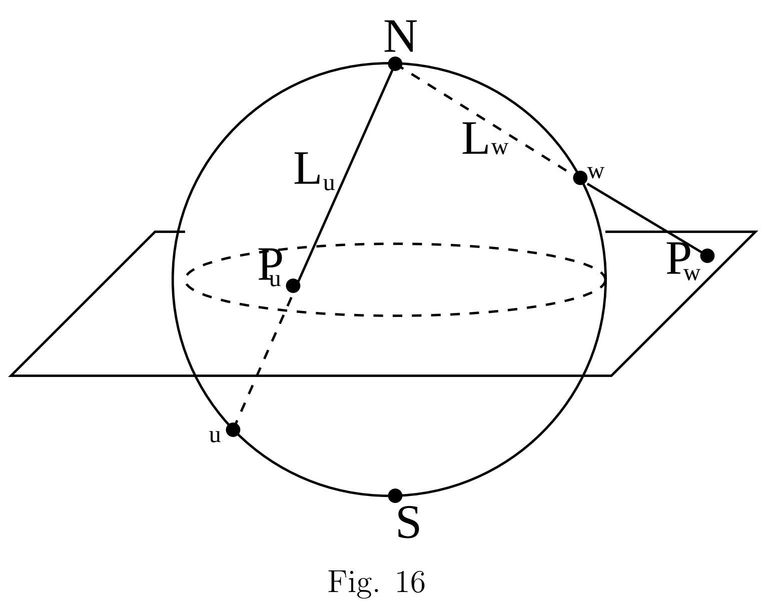


Fig. 16

Now, for any point $w \neq N$ on $\mathbb{S}^2$, let us denote by L_w, the line passing through the point w and N. This line intersects the (x_1, x_2)-plane in a unique point P_w determined as follows: if $w = (x_1, x_2, x_3)$, then the points of L_w are given by $(tx_1, tx_2, tx_3 + 1 - t), t \in \mathbb{R}$. Therefore, putting the last co-ordinate equal to zero, we get the point P_w in which L_w intersects the (x_1, x_2)-plane. It follows that $t = \dfrac{1}{1 - x_3}$ and hence

$$P_w = \left(\frac{x_1}{1 - x_3}, \frac{x_2}{1 - x_3}\right).$$

We now identify the (x_1, x_2)-plane in $\mathbb{R}^3$ itself with the complex plane, by mapping $(x_1, x_2, 0)$ to $z = x_1 + \imath x_2$. Thus, we obtain a continuous map $\sigma : \mathbb{S}^2 \setminus \{N\} \longrightarrow \mathbb{C}$ given by

$$(x_1, x_2, x_3) \mapsto \frac{x_1 + \imath x_2}{1 - x_3}.$$

This map is called the *stereographic projection.* It is easy to see, geometrically that σ is bijective: given any point $z = (y_1, y_2, 0)$, there is a unique point on $\mathbb{S}^2 \setminus \{N\}$ lying on the line L_z. Of course, we can even write down the formula for σ^{-1} :

$$z \mapsto \left(\frac{z + \bar{z}}{|z|^2 + 1}, \frac{z - \bar{z}}{\imath(|z|^2 + 1)}, \frac{|z|^2 - 1}{|z|^2 + 1}\right). \tag{3.36}$$

From this, it follows that σ^{-1} is also continuous. Such a map σ is called a *homeomorphism.*

Thus we have obtained a representation of $\mathbb{C}$ as the space of all unit vectors in $\mathbb{R}^3$ other than N. This is called the *spherical representation* of $\mathbb{C}$. Observe that under σ, the south pole S corresponds to the complex number 0 and every point on the unit circle goes to itself.

One easy fall-out of the spherical representation is that we now have another distance function on $\mathbb{C}$ coming from the usual 3–dimensional Euclidean distance. This can be given by the formula

$$\chi(z, z') = \frac{2|z - z'|}{[(1 + |z|^2)(1 + |z'|^2)]^{1/2}}. \tag{3.37}$$

(You will have to make good use of the cosine formula, in proving the above formula.) This is the so called *spherical metric* on $\widehat{\mathbb{C}}$. It is also called *Carathéodory*[5] *metric* or simply χ-*metric.* The following lemma is easy to prove(Ex.).

[5]Constantin Carathéodory(1873-1950) was German mathematician of Greek origin. He is famous for his studies in Calculus of variations, complex mappings and uniformization.

Lemma 3.8.1 A sequence $\{z_n\}$ of complex numbers tends to ∞ under the usual metric iff $\{\sigma^{-1}(z_n)\}$ tends to N under the χ-metric.

Because of this phenomenon, we can now make the following definition:

Definition 3.8.1 We call N the *point at infinity* for the complex plane. Often the complex plane $\mathbb{C}$ together with an extra point denoted by ∞ is denoted by $\widehat{\mathbb{C}}$. The mapping σ is then extended to $\sigma : \mathbb{S}^2 \longrightarrow \widehat{\mathbb{C}}$ by sending N to ∞.

This terminology is further justified by the following:

Lemma 3.8.2 The inversion map $z \mapsto \dfrac{1}{z}$ defined on $\mathbb{C} \setminus \{0\}$ extends to a unique homeomorphism of $\widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$ taking 0 to ∞ and vice versa.

Proof: The inversion map corresponds to the map

$$(x_1, x_2, x_3) \mapsto (x_1, -x_2, -x_3)$$

on $\mathbb{S}^2 \setminus \{N\}$ via σ. (Verify this.) This map is easily seen to extend uniquely to a homeomorphism which interchanges the north and the south pole. The conclusion of the lemma follows. ♠

Remark 3.8.1 * [Here, I am going to say things that are somewhat sophisticated and without proof. There is no need to panic even if you do not understand any of them, in your first reading. We do not use them in the subsequent material. However, do **not** skip this entirely.]

1. We can now extend the algebraic operations on $\mathbb{C}$ partially to cover the point at infinity also as follows:

$$\begin{array}{lcl} z + \infty & = & \infty \;=\; \infty + z \text{ for all } z \neq \infty \\ z.\infty & = & \infty \;=\; \infty.z \text{ for all } z \neq 0 \end{array}$$

$$\left.\begin{array}{lcl} z/0 & = & \infty, \\ z/\infty & = & 0 \end{array}\right\} \text{ for all } z \neq 0, \infty.$$

Experience tells us that these operations come quite handy often, and classically, that was the motivation to introduce the notion of extended complex plane. For instance, observe that if $z_n \longrightarrow z \neq \infty$, and $w_n \longrightarrow \infty$, then $z_n + w_n \longrightarrow \infty$ so that the sum formula for the limit is valid. Likewise, one can verify other limit rules also, extended in this sense. Observe that, it is not possible to define other operations such as $\infty + \infty$ etc., meaningfully, without getting into contradictory results. In the modern set up, there are several aspects of the extended complex plane and the algebraic motivation is the weakest of them all.

2. The spherical representation $\sigma : \mathbb{S}^2 \longrightarrow \widehat{\mathbb{C}}$ immediately tells us that the point at infinity is not merely 'hanging out' there. It has a neighborhood which looks like a neighborhood of any other point in the complex plane. The extension of inversion as in lemma 3.8.2, has actually established this relation with 0 and infinity. You can take the map $z \mapsto (z - z_0)^{-1}$ and see the same sort of relation between ∞ and any arbitrary point z_0. For example, lemma 3.8.1 tells us how to define the concept of continuous functions on the whole of $\widehat{\mathbb{C}}$. A function $f : \widehat{\mathbb{C}} \to X$ where X is any space, is continuous iff $f \circ \sigma$ is continuous. This is what is meant by defining a '*topology* on $\widehat{\mathbb{C}}$; the usual topology on the 2-sphere is transferred onto $\widehat{\mathbb{C}}$ via σ. (A subset $A \subset \widehat{\mathbb{C}}$ is open iff $\sigma^{-1}(A)$ is open in $\mathbb{S}^2$.) Moreover, it is now possible to define the concept of complex differentiability of a function f at ∞ defined in a neighborhood of ∞ : Call f differentiable at ∞ if the function $z \mapsto f(1/z)$ is differentiable at 0. Also, suppose Ω is an open subset of $\widehat{\mathbb{C}}$, $g : \Omega \longrightarrow \widehat{\mathbb{C}}$ and $z_0 \in \Omega$ such that $g(z_0) = \infty$. We say g is differentiable at z_0 if $1/g$ is differentiable at z_0. Essentially, this is what makes $\widehat{\mathbb{C}}$ into a *complex manifold.* (Observe that treated as a subspace of $\mathbb{R}^3$, the sphere does not inherit any such structure.) Together with this structure the extended complex plane is called the *Riemann Sphere.* This is the first non trivial example of a *Riemann Surface*, a connected 1-dimensional complex manifold. These were introduced by Riemann, in order to facilitate the study of complex functions, which are often multi-valued.

3. Here is then the formal definition of a Riemann surface. Let X be a topological space covered by open sets U_α such that to each α there is a homeomorphism $\phi_\alpha :$ $U_\alpha :\longrightarrow V_\alpha$ where V_α is an open subset of $\mathbb{C}$. Further suppose whenever $U_\alpha \cap U_\beta \neq \emptyset$ the map

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$$

is a biholomorphic mapping. We then call X a complex manifold of dimension 1 or equivalently a Riemann surface. Often we have to put additional conditions that X is a Hausdorff space and has a countable basis. Clearly all open subsets of $\mathbb{C}$ are trivially Riemann surfaces. $\widehat{\mathbb{C}}$ is the first nontrivial Riemann surface which is not an open subset of $\mathbb{C}$. We can cover it with two open sets $U_1 = \mathbb{C}$ and $U_2 = \mathbb{C} \setminus \{\infty\}$. We can then take $\phi_1 = Id$ and $\phi_2 : U_2 \to \mathbb{C}$ given by $\phi_2(z) = 1/z$. On the intersection of these two open sets viz. $\mathbb{C}^*$, we have $\phi_2 \circ \phi_1^{-1}(z) = 1/z$ which is a biholomorphic mapping.

4. Another important aspect of the extended complex plane is projective geometric and is a corner stone in Algebraic Geometry. Consider the set of all ordered pairs (z_1, z_2) of complex numbers wherein at least one of z_j is not equal to 0. Define an equivalence relation on this set as follows;

$$(z_1, z_2) \sim (w_1, w_2) \text{ iff there exists } \lambda \in \mathbb{C}^\star : (z_1, z_2) = \lambda(w_1, w_2).$$

   Denote by $[z_1 : z_2]$ the equivalence class represented by (z_1, z_2) and the set of equivalence classes by $\mathbf{P^1}(\mathbb{C})$. This is called the one-dimensional complex projective space.

   Observe that given a complex number z, we can identify it with the class $[z : 1]$. This will fill up all of $\mathbf{P^1}(\mathbb{C})$ except one point viz. $[1 : 0]$. For, any point (z_1, z_2) with $z_2 \neq 0$ is equivalent to $(z_1/z_2, 1)$ and all points $(z_1, 0)$ are equivalent to $(1, 0)$. We can easily identify this point with ∞ thereby completing the picture of the Riemann sphere.

   Also, we can interpret each class $[z_1 : z_2]$ to represent the complex 1-dimensional vector subspace of $\mathbb{C}^2$ spanned by the nonzero vector (z_1, z_2). Now, the true nature of the fractional linear transformations becomes visible. Let us start with a fractional linear transformation given by a 2×2 complex matrix A with non vanishing determinant. Then A defines a linear isomorphism of $\mathbb{C}^2 \longrightarrow \mathbb{C}^2$. In turn, this isomorphism defines a bijective mapping of the space of all complex 1-dimensional subspaces of $\mathbb{C}^2$ to itself. Under the above identification, it can now be verified that this map corresponds to the fractional linear transformation associated to A.

5. We can think of the extended complex plane as the so called '*one-point compactification* of $\mathbb{C}$. Indeed, the key for this is the lemma 3.8.1 already considered. The fact that $\mathbb{R}^3$ is a complete metric space and the closure of $\mathbb{S}^2 \setminus \{N\}$ is $\mathbb{S}^2$ in $\mathbb{R}^3$, can be used to see that the completion of $\mathbb{C}$ under the spherical metric is the extended complex plane.

6. The last but not the least important aspect is that $\mathbb{S}^2 \setminus N$ is a *conformal model* of the complex plane. By this, we just mean that the stereo-graphic projection σ is an angle preserving map. (See Ex. 19 and 20 in the Misc. Exercises to chapter 3 below.)

   All these aspects are important for a true understanding of the extended complex plane.

Exercise 3.8

1. Verify the formula (3.36) for σ^{-1}.

2. Prove the formula (3.37).

3. Compute $\chi(z, \infty)$.

4. Prove lemma 3.8.1.

5. Write down full details of the proof of theorem 3.7.4.

6. Show that stereographic projection establishes a one-to-one correspondence of the set of all circles on $\mathbb{S}^2$ with the set of all circles and straight lines on the complex plane. Also show that parallel lines in the plane will have their image circles which meet tangentially at the north pole on the sphere.

7. Determine all fractional linear transformations which map ∞ to ∞.

8. Given three distinct points $a, b, c \in \widehat{\mathbb{C}}$ determine all flts $\widehat{\mathbb{C}} \setminus \{a, b, c\} \to \mathbb{C} \setminus \{0, 1\}$. (See Exercise 3.7.5.)

9.* Show that every extended fractional linear transformation $T : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$ is a biholomorphic mapping of $\widehat{\mathbb{C}}$ to $\widehat{\mathbb{C}}$, (See remark 3.8.1.2)

3.9 Miscellaneous Exercises to Ch. 3

1. Let p be a polynomial function. Is the map $z \mapsto \dfrac{1}{p(z)}$ angle-preserving at points where it is defined?

2. Given two sets of distinct points $\{z_1, z_2, z_3\}, \{w_1, w_2, w_3\}$ show that the matrix

$$\begin{bmatrix} w_1 z_1 & w_1 & z_1 & 1 \\ w_2 z_2 & w_2 & z_2 & 1 \\ w_3 z_3 & w_3 & z_3 & 1 \end{bmatrix}$$

is of rank 3.

3. Given any four distinct points $z_j, \quad j = 1, 2, 3, 4$, show that there exists a fractional linear transformation which maps them respectively to $1, -1, w, -w$, where the choice of w depends on the points z_j. Find all possible values of w.

4. Show that the map

$$z \mapsto \frac{1}{2}\left(z + \frac{1}{z}\right)$$

defines a univalent transformation of the portion of $\boldsymbol{H}$ outside the unit circle onto $\boldsymbol{H}$. What is the image of $\mathbb{C} \setminus \mathbb{D}$ under this mapping? (This *quadratic* is called the *Joukowski function,* after the Russian mathematician who made it famous in aerodynamics.) In the figure below, the circle with center at $(x, 0)$ where $0 < x < 1$ and radius $1 + x$ is mapped onto the foil which has a cusp at the point -2.

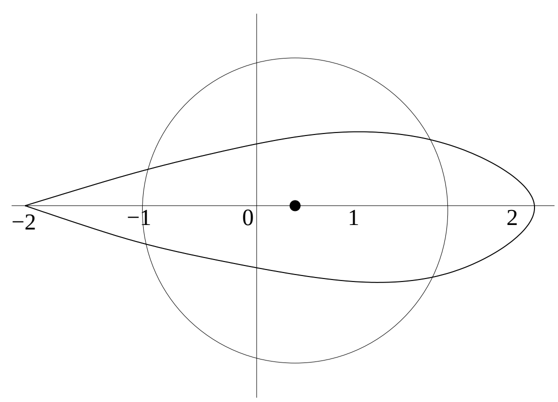


Fig. 17

5. Study the geometric properties of the map which is the composite of

$$z \mapsto \frac{z - z_0}{z - \bar{z}_0}; \quad z \mapsto \frac{z + 1}{z - 1}.$$

6. Examine the image of various lines in $\mathbb{C}$ (especially the lines $x = a$, and $y = b$) under the mapping $z \mapsto z^2$. Also, determine the image of the set $\{z \ : \ \Re(z) > 0\}$.

7. Describe the geometric nature of the families of curves given by the following equations for $-\infty < \lambda < \infty$:
   (a) $\Re\left(\frac{1}{z}\right) = \lambda$; (b) $\Im\left(\frac{1}{z}\right) = \lambda$; (c) $\Re z^2 = \lambda$; (d) $\Im z^2 = \lambda$.

8. For any two complex numbers b, c and $\lambda > 0$, determine the locus of the points z satisfying
$$|z^2 + bz + c| = \lambda.$$

9. Show that the mapping $w = \sin^2 z$ defines a biholomorphic mapping of the semi-infinite strip $\{(x, y) \ : \ 0 \leq x \leq \pi/2, y \geq 0\}$ onto the upper-half plane.

10. Obtain a bijective holomorphic mapping of the first quadrant Q_1 onto $\mathbb{D} \setminus (-1, 0]$.

11.* Obtain a biholomorphic mapping (preferably a fractional linear transformation) of $\boldsymbol{H}$ onto the portion of $\mathbb{C}$ lying below the parabola $y = x^2$. Try to get one onto the upper portion as well. Did you succeed in both cases? Did you get a fractional linear transformation in any of these cases? Compare your answer with Ex.15 below, before seeing the answer at the end of the book.

12. Show that the map
$$z \mapsto \cosh z$$
defines a biholomorphic equivalence of the half-strip $\{z \ : \ \Re z > 0, \ 0 < \Im z < \pi\}$ onto $\boldsymbol{H}$.

13. Let $A = \{x + \imath y \ : \ y^2 \leq \frac{1}{4} - x\}$. Show that $z \mapsto z^2$ defines a biholomorphic mapping of the infinite vertical strip $0 < \Re z < 1/2$ onto $A \setminus (-\infty, 0]$.

14. Let D be a region in $\mathbb{C}$ which is symmetric with respect to the antipodal action, i.e., $z \in D$ iff $-z \in D$. Assume that $0 \notin D$. Let $D_j, j = 1, 2$ be any two regions in $\mathbb{C}$ and $f_i : D \longrightarrow D_i, \ i = 1, 2$, be surjective holomorphic maps such that $f_i(w_1) = f_i(w_2)$ iff $w_1 = \pm w_2$. Then prove that there exists a biholomorphic mapping $\phi : D_1 \longrightarrow D_2$ such that $\phi \circ f_1 = f_2$.

15. Show that the mapping
$$z \mapsto z + \imath z^2$$
defines a biholomorphic mapping of the portion below the parabola $y = x^2$ onto $\boldsymbol{H}$. Next show that the map
$$z \mapsto \imath \cos \pi\sqrt{z}$$

maps the portion of $\mathbb{C}$ left to the parabola $y^2 = 1/4 - x$ biholomorphically onto the upper-half plane $\boldsymbol{H}$. (Observe that even though $\sqrt{z}$ is not well defined in the said region, the above map is well defined.) Use this to answer latter part of ex.11

16. Let C_1 and C_2 be any two circles on the sphere $\mathbb{S}^2$ intersecting at two points P_1 and P_2. Show that the angles of intersection at these two points are the same. (Call this angle α.)[Hint:Use a suitable reflection.]

17. Let $\sigma : \mathbb{S}^2 \longrightarrow \mathbb{R}^2$ denote the stereographic projection. Let $P_1 = N$ in the above exercise. Let $P' = \sigma(P_2)$. Then show that the two straight lines $\sigma(C_i)$ meet at P' at the same angle α as above.

18. Given any point $P_2 \neq N$ on $\mathbb{S}^2$ and a tangent line L to $\mathbb{S}^2$ through P_2, show that there exist a circle C passing through N and P_2 such that the line L is tangent to the circle C at P_2.

19. Show that σ is angle preserving on $\mathbb{S}^2 \setminus N$. [Hint: Use some of the above exercises.]

20.* Recall that the tangent space T_p to the sphere $\mathbb{S}^2$ at a point $P \in \mathbb{S}^2$ consists of all vectors in $\mathbb{R}^3$ perpendicular to the position vector OP.
(a) Establish the fact that any differentiable map $f : \mathbb{S}^2 \setminus N \longrightarrow \mathbb{R}^2$ is angle preserving iff $(Df)_p : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ restricted to T_p is angle preserving for all $p \in \mathbb{S}^2 \setminus N$.
(b) Using (a), prove that σ is angle preserving. [Now you have two proofs of the fact that σ is angle-preserving (the other one via ex. 18 and 19). Hold a friendly debate amongst yourselves on the merits and de-merits of the two proofs.]

21.* A *rhumb line* or a *loxodrome* is a smooth curve in $\mathbb{S}^2$ which makes a constant angle with all the meridian lines. How does the image of a rhumb line look like under σ?

22.* Consider the spherical co-ordinates for the unit sphere:

$$\xi = \cos\phi\cos\theta;\ \eta = \cos\phi\sin\theta;\ \zeta = \sin\phi;$$

$$-\frac{\pi}{2} \leq \phi < \frac{\pi}{2}; \quad -\pi \leq \theta < \pi.$$

Think of this as a mapping $f_1 : [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\pi, \pi] \rightarrow \mathbb{S}^2$. Is it angle preserving? (After you have thought well about it, hold an experiment by putting this question to a few of your friends amongst classmates, seniors and friendly teachers.)

23.* How does the inverse image $f_1^{-1}(L)$ of a rhumb line L look like under the spherical coordinate mapping f_1 as given in the above exercise.

24.* *The Mercator Mapping*: Here is an important application of the logarithm in navigation. Let $f_2 = \ln \sigma$ be the principle branch of logarithm of the stereographic projection, (well defined if we delete the *date-line* from the sphere. Is f_2 angle preserving? Let now $f_2 \circ f_1(\phi, \theta) = (u, v)$, where f_1 is the polar mapping as in ex. 22 above. Then show that $u = \ln(\sec \phi + \tan \phi)$ and $v = \theta$. Show that the image of a rhumb line under f_2 is a straight line. (Thus the navigators job of holding a constant course due north becomes simple as he has to only to follow a straight path on the Mercator map f_2.)

Chapter 4

Contour Integration

4.1 Definition and Basic Properties

The notion of a path as given in 1.6 is too general for our purpose here. From now on, we shall consider only a special class of paths which we shall call contours. We shall assume that you are familiar with basic theory of Riemann integration of a continuous function on a closed interval.

Definition 4.1.1 Let $\gamma : [a, b] \to \mathbb{R}^2$ be a path. We call γ a *contour* if it is *piecewise continuously differentiable,* i.e., if there exists a subdivision $a = a_0 < a_1 \cdots < a_n = b$ of the interval $[a, b]$, with the property that the restrictions $\gamma|_{[a_i,a_{i+1}]}$ are all continuously differentiable for each $i = 0, 1, 2, \ldots, n-1$.

It may be noted that a function defined on a closed interval is said to be differentiable on the closed interval if it is the restriction of a differentiable function defined on an open interval containing the closed interval.

Other notions such as closed contours, simple closed contours etc., are defined exactly in a similar fashion as in definition 1.6.1

Remark 4.1.1 Often, we confuse the image of a contour for the contour itself. This is good as far as there is no scope for further confusion, clear from the context such as in the expression 'w is a point on the contour γ' or 'let γ be a contour not passing through w' etc.. However, it should be noted that several contours may have the same image set and while performing integration on them, each of them may give different result. So, it is important to make the distinction between a function that defines a contour from the image set of a contour.

A typical example is $\gamma_k(t) = (\cos kt, \sin kt), \quad 0 \leq t \leq 2\pi$. For different integer values of k, these curves all have the same image $x^2 + y^2 = 1$. However each curve has different analytic and geometric behaviour as we shall soon see.

Example 4.1.1

1. **The line segment** Given any two points $z_1, z_2 \in \mathbb{C}$, the line segment from z_1 to z_2 is one of the simplest useful contour. We shall fix a parameterization of this viz.,

$$t \mapsto (1-t)z_1 + tz_2, \quad 0 \leq t \leq 1$$

and use the simple notation $[z_1, z_2]$ to denote it.

2. **The boundary of a rectangle** Let $a < b$ and $c < d$ be any real numbers. Consider the rectangle

$$R = \{(x, y) \ : \ a \leq x \leq b, \ c \leq y \leq d.\}$$

By the *boundary* ∂R (read as *daba* R) of this rectangle, we shall mean the contour obtained by tracing the line segments from (a, c) to (b, c) then from (b, c) to (b, d) then from (b, d) to (a, d) and finally from (a, d) to (a, b) in that order. It is easy to see that ∂R is a closed contour and its length is $2(b-a) + 2(d-c)$. The sense in which we have traced this contour is also referred to as 'counter-clock-wise.' Note that the image of this contour is nothing but δR (see definition 1.6.6) and that is the reason why we call this contour *boundary* of R.

3. **The boundary of a triangle** We shall denote the triangle with vertices a, b, c, by $T = \Delta(a, b, c)$. This is the set of all points $ta + sb + rc \in \mathbb{C}$ such that $t + s + r = 1, 0 \leq t, s, r \leq 1$. Assume that vertices have been labeled in the counterclockwise sense. The boundary ∂T is then the contour

$$\partial T = [a, b] \cdot [b, c] \cdot [c, a]$$

the composite of the three sides (traced in the counterclockwise sense). When we say a triangle T is contained in a set Ω we mean that this entire 2-dimensional object is contained in Ω and not just the boundary.

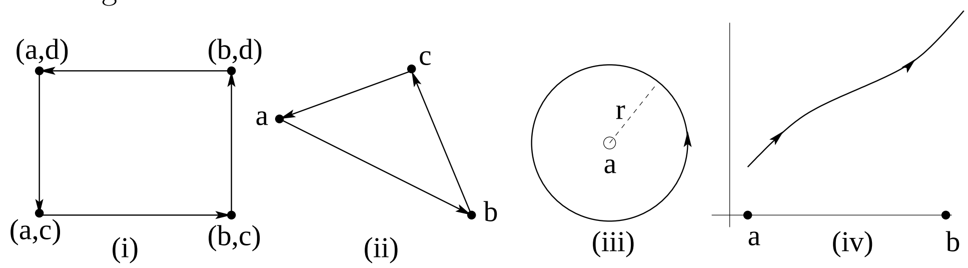


Fig. 18

4. **The boundary circle of a disc** Given an (open) disc $D = \{z \ : \ |z-a| < r\}$ of radius r and center a, the boundary $C = \partial D$ of this disc is the contour given by

$$t \mapsto a + e^{2\pi\imath\theta}, \quad 0 \leq \theta \leq 1.$$

   It should be noted that this contour is tracing the circle $|z-a| = r$ in the counterclockwise sense. However, often we may just express this contour by simply writing $|z-a| = r$.

5. **The graph** For any piecewise differentiable function $f : [a,b] \longrightarrow \mathbb{R}$, the graph of f is a contour parameterized by $t \mapsto (t, f(t))$.

6. The standard parameterizations of a circle or an ellipse, as seen earlier, are all examples of differentiable curves. An important class of contours is that of polygonal curves, viz., those consisting of finitely many straight line segments. Example 2 and 3 above were just two particular cases of this.

7. Consider the following function defined on $[0,1]$ by

$$f(t) = \begin{cases} 0, & t = 0 \\ 2nt - 1, & \dfrac{1}{2n} \leq t \leq \dfrac{1}{2n-1}, \ n \geq 1, \\ 1 - 2nt, & \dfrac{1}{2n+1} \leq t \leq \dfrac{1}{2n}, n \geq 1. \end{cases}$$

   Then f is continuous but the graph of f is not a contour. Equivalently observe that f is *not* piecewise differentiable. Draw the graph of f to see this! [Indeed, this is a standard example of a so called *non rectifiable curve,* i.e., the sum of the lengths of the line segments which make up the curve is infinite.]

Definition 4.1.2 By a *re-parameterization* of a smooth path $\gamma : [a,b] \longrightarrow A$, we mean a path $\gamma_1 : [c,d] \longrightarrow A$ such that there exists a continuously differentiable function $\tau : [c,d] \longrightarrow [a,b]$ with $\tau'(t) > 0$ for all $t \in (c,d)$ and $\gamma_1(t) = \gamma \circ \tau(t), \ c \leq t \leq d$. Observe that $Im(\gamma) = Im(\gamma_1)$. By a *re-parameterization* of a contour we mean a path which restricts to differentiable reparameterization of each of the differentiable pieces of the contour.

At this stage, we assume that you are familiar with the concept of Riemann integration of a continuous real valued function on a closed interval. Suppose now that $f : [a, b] \longrightarrow \mathbb{C}$ is continuous. Then we define

$$\boxed{\int_a^b f(t)\,dt := \int_a^b \Re(f(t))\,dt + \imath \int_a^b \Im(f(t))\,dt.} \tag{4.1}$$

Standard properties Riemann integrals of real valued continuous functions all hold for the above complex valued integrals also. Linearity properties are easy to check. However, properties involving inequalities need to be checked carefully. For instance, consider the following familiar inequality

$$\boxed{\left|\int_a^b f(t)\,dt\right| \leq \int_a^b |f(t)|dt.} \tag{4.2}$$

Try to prove this yourself and then read the proof below.
Proof: Express the integral on the left as $re^{i\theta}$. Then

$$\left|\int_a^b f(t)\,dt\right| = r = e^{-i\theta}\int_a^b f(t)\,dt = \int_a^b e^{-i\theta} f(t)\,dt.$$

Upon taking the real parts of both sides, we get,

$$r = \int_a^b \text{ Re } (e^{-i\theta} f(t))\,dt \leq \int_a^b |f(t)|dt.$$

Definition 4.1.3 Let $\omega : [a, b] \longrightarrow \mathbb{C}$ be a path such that ω' is continuous. Then for any continuous function f defined on the image of ω, we define the *contour integral* or the *line integral* of f along ω to be

$$\boxed{\int_\omega f\,dz := \int_a^b f(\omega(t))\omega'(t)\,dt.} \tag{4.3}$$

Here the prime "′" denotes differentiation with respect to t. Observe that $\omega'(t)$ is a complex number for each t, say, $\omega(t) = x(t) + \imath y(t)$, then $\omega'(t) = x'(t) + \imath y'(t)$. Similarly if we write $f(z) = u(z) + \imath v(z)$, then $f(\omega(t)) = u(\omega(t)) + \imath v(\omega(t))$. Hence the R.H.S. of the above definition can also be expressed as

$$\int_a^b (u(\omega(t))x'(t) - v(\omega(t))y'(t))\,dt + \imath \int_a^b (u(\omega(t))y'(t) + v(\omega(t))x'(t))\,dt.$$

In calculus of two real variables, this is written in the form

$$\left(\int_\omega udx - vdy, \;\; \int_\omega udy + vdx\right).$$

Remark 4.1.2 Observe that these definitions are equivalent to introducing the formal symbols $dt,\ dx,\ dy,\ dz$ etc. by the formulae:

$$dx := x'(t)\,dt;\;\; dy := y'(t)\,dt;\;\; dz := \omega'(t)\,dt = (x'(t) + \imath y'(t))\,dt = dx + \imath dy.$$

These symbols can be multiplied by continuous functions to obtain other symbols such as $f\,dx,\ g\,dz$ etc. Further any two such symbols can be added together, to get what one generally calls a *1-differential*, or a *differential 1-form.* For our purpose, viz., for the study of the contour-integration, it is enough to know the linearity properties of these symbols, viz.,

$$\boxed{\boxed{(f+g)\,dx = f\,dx + g\,dx; \quad (\alpha f)\,dx = \alpha(f\,dx).}} \tag{4.4}$$

Example 4.1.2 Let us compute the value of $\displaystyle\int_\omega x\,dz$, where ω is the line segment from 0 to $1+\imath$. We can choose any parameterization of ω, say, $\omega(t) = (1+\imath)t, \quad 0 \le t \le 1.$ Then $\omega'(t) = 1+\imath$ for all t and hence by definition

$$\int_\omega x\,dz = \int_0^1 x(\omega(t))\omega'(t)\,dt = \int_0^1 t(1+\imath)\,dt = (1+\imath)/2.$$

Example 4.1.3 Let us compute $\displaystyle\int_{|z-a|=r} (z-a)^n dz$ where n is a given integer. Then By definition, this integral is equal to

$$\int_0^{2\pi} r^n e^{n\imath\theta} d(r^n e^{\imath\theta}) = \int_0^{2\pi} \imath e^{(n+1)\imath\theta} d\theta.$$

Therefore, for $n \neq -1$ we have

$$\int_{|z-a|=r} (z-a)^n dz = 0$$

whereas for $n = -1$ we have

$$\boxed{\int_{|z-a|=r} \frac{1}{z-a}\, dz = 2\pi\imath.} \tag{4.5}$$

More generally, if f is a complex differentiable function in a domain Ω and γ is a closed contour in Ω, let us put $g(t) = f(\gamma(t), a \leq t \leq b$. By chain rule we have, $g'(t) = f'(\gamma(t))\gamma'(t)$. Therefore,

$$\int_\gamma f'(z)dz = \int_a^b f'(\gamma(t))\gamma'(t)dt = \int_a^b g'(t)dt = g(b) - g(a) = 0,$$

since $\gamma(b) = \gamma(a)$. The situation in the previous example is similar: for $n \neq -1$, the integrand is the derivative of a function, and the computation for $n = -1$ now proves the fact that the integrand in this case **cannot** be the derivative of a function on an open disc around 0.

Remark 4.1.3 If $f(z) = \sum_n a_n(z-z_0)^n$ is given by a convergent power series, term-by-term integration and the above computation shows that $\int_{|z-z_0|=r} f(z)dz = 0$ for all $0 \leq r < R$, where R is the radius of convergence. We shall prove such a result soon for all complex differentiable functions, and this is going to be the central result in contour integration.

Remark 4.1.4 Some basic properties of the integral: Here we list a number fundamental properties of contour integrals which are easy consequences of corresponding properties of Riemann integrals, of which the reader is supposed to be familiar. She may look into any elementary book on real analysis for more details. (See for example [Ru-1].)

1. **Change of Parameterization**

   The most basic property of our integral $\displaystyle\int_\gamma f(z)\, dz$ is the invariance under change of parameterization. So, let $\tau : [\alpha, \beta] \longrightarrow [a, b]$ be a continuously differentiable function with $\tau(\alpha) = a,\ \tau(\beta) = b,\ \tau'(t) > 0,\ \forall\ t$. Then

$$\boxed{\int_{\gamma\circ\tau} f(z)\, dz := \int_\alpha^\beta f(\gamma\circ\tau(t))\frac{d(\gamma\circ\tau)}{dt}(t)\, dt.} \tag{4.6}$$

   By chain rule $\frac{d(\gamma\circ\tau)}{dt}(t) = \gamma'(\tau(t))\tau'(t)$. Putting $s = \tau(t)$ and hence $ds = \tau'(t)\, dt$, it follows that the R.H.S. is equal to

$$\int_a^b f(\gamma(s))\gamma'(s)\, ds = \int_\gamma f(z)\, dz.$$

Therefore,

$$\int_{\gamma\circ\tau} f(z)\,dz = \int_{\gamma} f(z)\,dz \tag{4.7}$$

2. **Linearity**

   The usual linearity properties of the integral are all valid here, viz., for all $\alpha,\ \beta \in \mathbb{C}$

$$\boxed{\int_{\gamma}(\alpha f + \beta g)(z)\,dz = \alpha\int_{\gamma} f(z)\,dz + \beta\int_{\gamma} g(z)\,dz.} \tag{4.8}$$

3. **Additivity Under Sub-division** or **Concatenation**

   If $a < c < b$ and $\gamma_1 = \gamma|_{[a,c]},\ \gamma_2 = \gamma|_{[c,b]}$, are the restrictions to the respective sub-intervals then

$$\boxed{\int_{\gamma_1\cdot\gamma_2} f(z)\,dz := \int_{\gamma_1} f(z)\,dz + \int_{\gamma_2} f(z)\,dz = \int_{\gamma} f(z)\,dz.} \tag{4.9}$$

4. **Orientation Respecting** We also have,

$$\boxed{\int_{\gamma^{-1}} f(z)\,dz = -\int_{\gamma} f(z)\,dz} \tag{4.10}$$

   where γ^{-1} is the curve γ itself traced in the opposite direction, viz., $\gamma^{-1}(t) = \gamma(a+b-t)$. To see this, put $t = a+b-s$. Then,

$$\begin{aligned} L.H.S. &= \int_a^b f(\gamma^{-1}(s))\frac{d\gamma^{-1}}{ds}(s)\,ds \\ &= \int_b^a f(\gamma(t))\gamma'(t)(-dt) \\ &= -\int_a^b f(z)\,dz = R.H.S. \end{aligned}$$

   For this reason, we could also use the notation $-\gamma$ for γ^{-1}.

5. **Interchange of order of integration and limit**

   If $\{f_n\}$ is a sequence of continuous functions **uniformly converging** to f then the limit and integration can be interchanged viz.,

$$\boxed{\lim_{n\to\infty}\int_{\gamma} f_n(z)\,dz = \int_{\gamma} f(z)\,dz.} \tag{4.11}$$

6. **Interchange of order of iterated integration:** Suppose $f(z,w)$ is a continuous function of two variables, γ, ω are two contours, such that $\displaystyle\int_\gamma f(z,w)dz$ is defined for all values of the variable w and $\displaystyle\int_\omega f(z,w)dw$ is defined for all values of the variable z. Then the two iterated integrals are defined and are equal:

$$\boxed{\int_\gamma \left(\int_\omega f(z,w)dw\right) dz = \int_\omega \left(\int_\gamma f(z,w)dz\right) dw} \tag{4.12}$$

This follows directly from the so called Fubini's theorem for double intergals on rectangular region.

7. **Term-by-term Integration** From (5) it also follows that whenever we have a uniformly convergent series of functions then *term-by-term* integration is valid.

$$\boxed{\int_\gamma \left(\sum_n f_n(z)\right)\, dz = \sum_n \left(\int_\gamma f_n(z)\, dz\right).} \tag{4.13}$$

8. **Fundamental Theorem of Integral Calculus** Let γ be a (continuous) contour in a domain Ω and f be a holomorphic function on Ω. Then $\int_\gamma f'(z)dz = f(b)-f(a)$ where a and b are the initial and terminal points of γ respectively.

   This is a direct consequence of the corresponding result in 1-variable Riemann integration.

   We now extend our definition of the integral to cover contours also. Thus suppose that a contour ω is broken up into a number of differentiable arcs $\omega = \omega_1.\cdots.\omega_k$, we see that property (3) comes to our help and says that the only natural way to define the integrals over arbitrary contours is by the formula

$$\int_\omega f(z)\, dz := \sum_{j=1}^k \int_{\omega_j} f(z)\, dz.$$

   Verify directly that properties 1-5 are all valid in this generality as well.

   For future use let us introduce a notation here. First, for a differentiable curve ω, and a continuous function f on the image of ω taking real or complex values, put

$$\int_\omega |f(z)\, dz| \quad := \quad \int_a^b |f(\omega(t))\omega'(t)|\, dt. \tag{4.14}$$

More generally, for a contour $\omega = \omega_1.\cdots.\omega_k$, which is the composite of differentiable arcs $\omega_j = \omega|_{[a_j,a_{j+1},]}$ $a = a_1 < a_2 < \cdots < a_{k+1} = b$ and for any continuous function f on ω, we shall have the notation

$$\int_\omega |f(z)\,dz| := \sum_j \int_{\omega_j} |f(z)s| = \sum_{i=1}^k \int_{a_j}^{a_{j+1}} |f(\omega_j(t))\omega_j'(t)|\,dt. \tag{4.15}$$

Just as before we can verify that these quantities are independent of reparameterization. Indeed, the strong condition that $\tau'(t) > 0$, in the definition of the reparameterization, will be needed here, for the first time.

9. With f and ω as above, as a consequence of (4.2), we have,

$$\boxed{\left|\int_\omega f(z)dz\right| \le \int_\omega |f(z)dz|.} \tag{4.16}$$

10. The continuity assumption on the function f is quite strong. The entire discussion above is valid whenever the function is 'Riemann integrable' on the contour. Thus, for example, we can allow f to be discontinuous at some finitely many points of the contour γ and require it to be bounded, then the integral $\int_\gamma f(z)dz$ makes sense and has all the properties discussed above. This remark plays a crucial role in Cauchy integration theory later.

One special case the notation (4.14) corresponds to the geometric notion of *arc length:*

Definition 4.1.4 Length of a contour: *Let* $\omega : [a,b] \longrightarrow \mathbb{R}^2$, $\omega(t) = (x(t), y(t))$ *be a continuously differentiable arc. Then the* arc-length *of* ω *is obtained by the integral*

$$\boxed{L(\omega) = \int_a^b [((x'(t))^2 + (y'(t))^2]^{1/2}\,dt.} \tag{4.17}$$

Using change of variable formula for integrals on intervals, it follows that $L(\omega)$ is independent of the choice of parameterization of ω as discussed earlier. In complex notation, $\omega(t) = z(t) = x(t) + \imath y(t)$, this becomes

$$L(\omega) := \int_\omega |dz| \tag{4.18}$$

Thus, if you have any difficulty in understanding what the R.H.S. in (4.18) stands for, remember that this symbol stands for R.H.S. in (4.17). Note that the definition easily extends to all contours via (4.15).

Example 4.1.4 Let us compute the length of the circle $C_r := z(\theta) = re^{\imath\theta}, \;\; 0 \le \theta \le 2\pi$.

$$L(C_r) = \int_{C_r} |d(re^{\imath\theta})| = \int_0^{2\pi} (r^2\sin^2\theta + r^2\cos^2\theta)^{1/2} d\theta \; = \; r\int_0^{2\pi} d\theta = 2\pi r.$$

Theorem 4.1.1 M-L inequality *Let A be an open set in $\mathbb{C}$, f be a continuous function on A and $\omega : [a,b] \longrightarrow A$ be a contour in A. Let $M = sup\{|f(z)| \;:\; z \in Im(\omega)\}$. Then*

$$\left|\int_\omega f(z)\,dz\right| \;\le\; ML(\omega).$$

Proof: From (4.16) we have,
$$\left|\int_\omega f(z)\,dz\right| \le \int_\omega |f(z)\,dz| = \int_a^b |f(\omega(t))|\,|\omega'(t)|\,dt \le \;M\int_a^b |\omega'(t)|\,dt = ML(\omega). \qquad \spadesuit$$

As an immediate corollary to M-L inequality, one can prove property (4.11) and then use this to prove (4.12) and (4.13).

We leave this as an exercise to you.

Theorem 4.1.2 *Let U be an open set in $\mathbb{R}^n$ or $\mathbb{C}^n$. Let $g : U \times [a,b] \longrightarrow \mathbb{C}$ be a continuous function and $\phi(P) = \displaystyle\int_a^b g(P,t)\,dt, \;\; P \in U$. Then $\phi : U \longrightarrow \mathbb{C}$ is a continuous function. In particular, if ω is a contour in $\mathbb{C}$ and $g_1 : U \times Im(\omega) \longrightarrow \mathbb{C}$ is a continuous function, then $\phi_1(P) := \displaystyle\int_\omega g_1(P,z)\,dz$ is continuous on U.*

Proof: Let B be a closed ball of radius, say $\delta_1 > 0$, around a point $P_0 \in U$ such that $B \subset U$. Then $B \times [a,b]$ is a closed and bounded subset of $\mathbb{C}^n \times \mathbb{C}$. Hence, g restricted to this set is uniformly continuous. This means that given $\epsilon > 0$, we can find a $\delta_2 > 0$ such that

$$|g(P_1,t_1) - g(P_2,,t_2)| < \frac{\epsilon}{(b-a)}$$

for all $(P_i,t_i) \in B \times [a,b]$ whenever $\|(P_1,t_1) - (P_2,t_2)\| < \delta_2$. Now let $\delta = min\{\delta_1,\delta_2\}$ and $|P - P_0| < \delta$. Then

$$|\phi(P) - \phi(P_0)| \;=\; \left|\int_a^b (g(P,t) - g(P_0,t))\,dt\right| \;\le\; \epsilon.$$

This proves the continuity of ϕ at P_0. Since P_0 is an arbitrary point of U, it follows that ϕ is continuous on U. The latter part follows by taking $g(P,t) = g_1(P,\omega(t))$. $\spadesuit$

Theorem 4.1.3 Differentiation Under the Integral Sign *Let U be an open subset of $\mathbb{C}$ and $g : U \times [a,b] \longrightarrow \mathbb{C}$ be a continuous functions such that for each $t \in [a,b]$, the function $z \mapsto g(z,t)$ is complex differentiable and the map $\dfrac{\partial g}{\partial z} : U \times [a,b] \longrightarrow \mathbb{C}$ is continuous. Then in U, the integrated function*

$$f(z) = \int_a^b g(z,t)dt$$

is complex differentiable and

$$f'(z) = \int_a^b \frac{\partial g}{\partial z}(z,t)\, dt.$$

Proof: Given $z_0 \in U$, let $r > 0$ be such that $B = \bar{B}_r(z_0) \subset U$. Then $B \times [a,b]$ is closed and bounded and hence $\dfrac{\partial g}{\partial z}$ is uniformly continuous on it. Hence, given $\epsilon > 0$ we can choose $0 < \delta < r$ such that

$$\left|\frac{\partial g}{\partial z}(z_1,t) - \frac{\partial g}{\partial z}(z_2,t)\right| < \frac{\epsilon}{b-a} \tag{4.19}$$

for all $t \in [a,b]$ and $z_1, z_2 \in B$ such that $|z_1 - z_2| < \delta$. Now, let $0 < |z - z_0| < \delta$. Then

$$\begin{aligned} \left|(g(z,t) - g(z_0,t) - (z-z_0)\tfrac{\partial g}{\partial z}(z_0,t)\right| &= \left|\int_{[z_0,z]} \left(\frac{\partial g}{\partial w}(w,t) - \frac{\partial g}{\partial z}(z_0,t)\right) dw\right| \\ &\leq \left|\int_{[z_0,z]} \frac{\epsilon}{b-a} dw\right| = \frac{\epsilon|z-z_0|}{b-a}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\left|\frac{f(z)-f(z_0)}{z-z_0} - \int_a^b \frac{\partial g}{\partial z}(z_0,t)\, dt\right| \\ = \;& \frac{1}{|z-z_0|}\left|\int_a^b (g(z,t) - g(z_0,t))\, dt - (z-z_0)\int_a^b \frac{\partial g}{\partial z}(z_0,t)\, dt\right| \\ = \;& \frac{1}{|z-z_0|}\left|\int_a^b \left[(g(z,t) - g(z_0,t) - (z-z_0)\frac{\partial g}{\partial z}(z_0,t)\right] dt\right| \leq \epsilon. \end{aligned}$$

This proves the theorem. ♠

Theorems 4.1.2, 4.1.3 are going to be extremely useful. Even though we have proved the result for complex differentiable functions, this holds for real differentiable functions as well, and the proof is the same. As a simple minded application of the second one let us prove:

Theorem 4.1.4 *For all points w such that $|w-a|<r$, we have*

$$\boxed{\boxed{\int_C \frac{dz}{z-w} = 2\pi\imath.}} \tag{4.20}$$

where C is the positively oriented boundary of the disc $|z-a|=r$.

Proof: We have already proved this result for $w=a$ in Example 4.1.3. Now fix any w and define $g:[0,1]\to\mathbb{C}$ by

$$g(t) = \int_C \frac{dz}{z-tw}.$$

Then by theorem 4.1.3, g is differentiable and its derivative can be computed by differentiating under the integral sign:

$$g'(t) = -w\int_C \frac{dz}{(z-tw)^2} = w\int_C \frac{d}{dz}\left(\frac{1}{z-tw}\right) = 0,$$

the last equality being a consequence of the fundamental theorem of integral calculus. Therefore, g is a constant and we have $g(1)=g(0)=2\pi\imath$, as required. ♠

Exercise 4.1

1. Find the length of the following curves:

   (i) The line segment joining 0 and $1+\imath$.

   (ii) The hypo-cycloid given by: $x=a\cos^3\theta,\ y=a\sin^3\theta,\quad 0\le\theta\le 2\pi$, where, $a>0$ is a fixed number.

   (iii)* The perimeter of an ellipse with major and minor axes of size a and b respectively. (*Caution: this is rather a difficult problem.*)

2. Compute $\displaystyle\int_{|z|=\rho} x\,dz$, where $|z|=\rho$ is the circle of radius ρ around 0 taken in the counter clockwise sense.

3. Compute $\displaystyle\int_{|z|=\rho} z^n\,dz$, for all integers n. Use this to compute the integral in Ex. 2 in a different way, by writing $x=(z+\bar{z})/2=(z+\rho^2z^{-1})/2$.

4. Suppose $f(z)$ is holomorphic in a domain containing a closed curve C. (The hypothesis about continuity of f' is redundant but we have not proved this yet.) Prove that $\displaystyle\int_C \overline{f(z)}f'(z)dz$ is purely imaginary.

5. Prove (4.11) and (4.12). [Hint: Use ML inequality.]

6. * Show that the curve in example 4.1.1.7 has *infinite length.*

4.2 Existence of Primitives

In this section, we shall study the question of finding a differentiable function g whose derivative is the given function f. Such a function g is called a *primitive* of f. Of course, in general, it may not even exist. When f is a function of a real variable, recall that this question is answered by merely taking the semi-indefinite integral of the given function. We have no reason not to follow this procedure even in the 2-variable case. So, let us begin.

Assume that f is a continuous complex valued function over a region Ω. We can then associate to each contour ω in Ω a real number $\displaystyle\int_\omega f dz$. In order to get a function on the domain Ω itself, we should first of all fix up the initial point for all these contours say, $z_0 \in \Omega$. Now given any point $z \in \Omega$, we may choose a contour ω from z_0 to z and consider the integral. We perceive another problem. The value of this integral depends on the choice of the contour joining z_0 to z. Thus we are led to consider only such functions f for which the integral is independent of the choice of ω joining z_0 to z. We shall soon see that this restriction is indeed quite reasonable and moreover, will yield the required result.

Example 4.2.1 Consider function $f(z) =$ the principle value of $\sqrt{z}$. Let us integrate this on the upper semi-circle:

$$C_1 : \theta \mapsto e^{i\theta}, \quad 0 \leq \theta \leq \pi.$$

By definition, we have

$$\int_{C_1} f(z)dz = \int_0^\pi e^{i\theta/2} d(e^{i\theta}) = -2(i+1)/3.$$

But on the lower semi-circle $C_2 : \theta \mapsto e^{-i\theta}, \quad 0 \leq \theta \leq \pi$, observe that $f(z)$ has a discontinuity at the end point π. Therefore, the integral has to be calculated carefully. We have,

$$\begin{aligned}\int_{C_2} f(z)dz &= \lim_{s\to\pi} \int_0^s e^{-i\theta/2} d(e^{-i\theta}) = \lim_{s\to\pi} \int_0^s -\imath e^{-3i\theta/2} d\theta \\ &= \lim_{s\to\pi} \frac{2}{3}(e^{-3is/2} - 1) = \frac{2}{3}(i-1).\end{aligned}$$

Thus the two integrals are different which shows that the integral is path-dependent. This phenomenon is explained by the fact that there is no anti-derivative of $\sqrt{z}$ in a domain which 'encircles' the origin. We shall make this phenomenon clearer in what follows.

Theorem 4.2.1 *Let Ω be a region in $\mathbb{C}$, and p, q be continuous maps on Ω taking real (or complex) values. Then the following conditions are equivalent.*

(a) *The differential $pdx + qdy$ is exact in Ω, i.e., there exists real (or complex) valued function u on Ω such that*

$$\frac{\partial u}{\partial x} = p \quad \textit{and} \quad \frac{\partial u}{\partial y} = q. \tag{4.21}$$

(b) *For all closed contours ω in Ω, we have,*

$$\int_\omega (pdx + qdy) = 0. \tag{4.22}$$

Proof: [By taking real and imaginary parts separately, the statement of the theorem for complex valued functions follows from that for real valued functions. Therefore, you can assume that only real valued functions appear in the proof below. However, such an assumption is not a logical necessity.]

Let $\omega : [a, b] \longrightarrow \Omega$ be a contour joining z_1 and z say, given by $\omega(t) = (x(t), y(t))$. Suppose $du = pdx + qdy$. Then by definition,

$$\begin{aligned}
\int_\omega (pdx + qdy) &= \int_a^b [p(\omega(t))x'(t) + q(\omega(t))y'(t)]\, dt \\
&= \int_a^b \left(\frac{\partial u}{\partial x} x'(t) + \frac{\partial u}{\partial y} y'(t) \right) dt \\
&= \int_a^b \frac{d}{dt}(u(x(t), y(t))\, dt = u(x(b), y(b)) - u(x(a), y(a)) \\
&= u(\omega(b)) - u(\omega(a)) = u(z) - u(z_0).
\end{aligned}$$

Observe that we have used the fundamental theorem of integral calculus of 1-variable above. Now, if ω is closed, then $z_0 = z$ and hence $\int_\omega (pdx + qdy) = 0$. This proves (a) $\Longrightarrow$ (b).

To prove (b) $\Longrightarrow$(a), fix any point $z_0 \in \Omega$. Then for every point $z \in \Omega$, choose a piecewise differentiable path γ_z from z_0 to z in Ω. Define

$$u(z) := \int_{\gamma_z} (pdx + qdy) \tag{4.23}$$

Let us proceed to prove that $du = pdx + qdy$, i.e., $\dfrac{\partial u}{\partial x} = p$, $\dfrac{\partial u}{\partial y} = q$. Given $z = (x, y) \in \Omega$, choose sufficiently small $\epsilon > 0$, so that $(x + h, y) \in \Omega$ for all $|h| < \epsilon$.

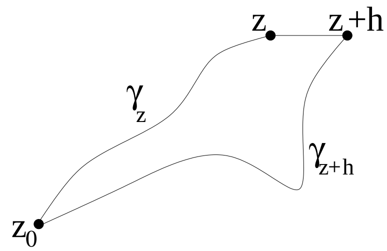


Fig. 19

Now restrict h further, to be a real number. We have two specific ways of approaching the point $z+h$ from z_0. One is along the chosen path γ_{z+h}. The other one is to first trace γ_z and then trace the line segment $[z, z+h]$. Condition (b) implies that

$$u(z+h) := \int_{\gamma_{z+h}} pdx + qdy = \int_{\gamma_z} pdx + qdy + \int_{[z,z+h]} pdx + qdy.$$

Therefore,

$$u(z+h) - u(z) = \int_{[z,z+h]} (pdx + qdy) \tag{4.24}$$

Now recall that the segment $[z, z+h]$ is parameterized by

$$t \mapsto (x+th, y), 0 \le t \le 1.$$

Therefore, $dx = hdt$ and $dy = 0$. Thus

$$u(z+h) - u(z) = \int_{[z,z+h]} (pdx + qdy) = \int_0^1 p(x+th, y)hdt = p(x+t_0h, y)h$$

for some $0 \le t_0 \le 1$, by the Mean Value Theorem of integral calculus of 1-real variable.

Now divide by h, take the limit as $h \longrightarrow 0$, and appeal to the fact that p is continuous to get, $\dfrac{\partial u}{\partial x}(x, y) = p(x, y)$.

The proof that $\dfrac{\partial u}{\partial y} = q$ is similar, by taking $\imath h$ in place of h. ♠

Corollary 4.2.1 In the situation of theorem 4.2.1, assume further that Ω is a convex region. Then (a), (b) are equivalent to the following:
(c) For all triangles T contained in Ω

$$\int_{\partial T} pdx + qdy = 0. \tag{4.25}$$

Proof: The implication (b) $\Longrightarrow$ (c) is obvious. To prove (c) $\Longrightarrow$ (a) we imitate the proof of (b) $\Longrightarrow$ (a) except that we now take γ_z to be the line segment $[z_0, z]$ from z_0 to z. (This is where convexity of Ω is used.) Then the hypothesis (c) is enough to arrive at (4.24) since the closed path $\gamma_z \star [z, z+h] \star \gamma_{z+h}^{-1} = \partial T$ is the boundary of a triangle $T = \Delta(z_0, z, z+h)$ in Ω. The rest of the proof is as before. ♠

Remark 4.2.1 The function u in the above theorem, if it exists, is unique up to an additive constant. (Why?) The ambiguity in the additive constant is a cheap price we pay for the freedom we enjoy in the choice of the base point z_0.

Corollary 4.2.2 Primitive Existence Theorem: For a continuous complex valued function f defined in a region Ω, the integral $\int_\omega f\,dz = 0$ for all closed contours ω iff f is the derivative of a holomorphic function on Ω.

Proof: Suppose there is a holomorphic function g such that $g' = f$. By CR equations, we have $f = g_x = -\imath g_y$ and hence

$$f(z)dz = f(z)(dx + \imath dy) = g_x dx + g_y dy.$$

Therefore from the above theorem, it follows that, $\int_\omega f\,dz = 0$ for all closed contours in Ω. Conversely, suppose $\int_\omega f\,dz = 0$ for all closed contours in Ω then by taking $p = f$ and $q = \imath f$ in the above theorem, it follows that there exists $F : \Omega \longrightarrow \mathbb{C}$, such that

$$\frac{\partial F}{\partial x} = f; \quad \frac{\partial F}{\partial y} = \imath f.$$

This implies that F satisfies the CR equations: $F_x + \imath F_y = 0$. Since, f is continuous, the partial derivatives of F are continuous. Therefore F is complex differentiable and $F'(z) = F_x = f$. This completes the proof of the corollary. ♠

Remark 4.2.2 In the next section, we shall see that complex differentiable functions over nice domains satisfy the condition required by the above theorem. This is what is known as Cauchy's theory. Its importance in the theory of complex functions cannot be over-emphasized.

Example 4.2.2 As seen in the example 4.1.3, in the previous section, $\displaystyle\int_{|z-a|=r} \frac{dz}{(z-a)} \neq 0$. It follows that $\dfrac{1}{z-a}$ does not have a primitive in any neighborhood of $z = a$. Equivalently, this means that we cannot define $log\,(z-a)$ in any neighborhood of $z = a$, as a single valued function. (Of course, in a small neighborhood of any other point, it is the derivative of a holomorphic function.)

In fact, as we shall see later, it is not possible to do this even in any region that contains an annulus A :

$$\rho_1 < |z-a| < \rho_2.$$

Here, $0 \leq \rho_1 < \rho_2$. We shall also prove that there are no well defined branches of $\sqrt[n]{z}, n \geq 2$, in A. These are negative results; we shall also see some positive results about defining log as a single valued function.

4.3 Cauchy-Goursat Theorem

There are mainly two aspects of Cauchy's theorem. The analytic aspect is deeper but easy to state and needs less background to understand. The topological aspect is technical in nature but is necessary to make the result more applicable. This will also help bring out the true nature of the result. Accordingly, there are several forms of this theorem. We shall build it up slowly, postponing delicate, time consuming but rather peripheral issues to a latter chapter and concentrating on the central theme here.

One of the most striking features of Cauchy's theory is in bringing out the strength of the complex differentiability. This is enhanced by the following theorem, the basic idea of the proof of which is due to E. Goursat.[1] Indeed, this gives a quick proof of Cauchy's theorem and integral formula with no **'extra'** assumptions on the function f other than complex differentiability.

You may choose to skip this section and instead go through the next section. The only price you pay is that you will not have seen a proof of complex differentiability implies holomorphicity. The best thing would of course will be to study both sections.

Theorem 4.3.1 Cauchy-Goursat Theorem on a Triangle: *Let f be a complex differentiable function in a region Ω. Let $T = \Delta(a, b, c)$ be a triangle contained in Ω and let ∂T denote the contour obtained by traversing its boundary in the counter clockwise*

[1]Edouard Goursat(1858-1936), a French mathematician.

direction. Then

$$\int_{\partial T} f(z)\,dz = 0.$$

Proof: Let us introduce the notation

$$s(T_0) = \int_{\partial T_0} f(z)\,dz$$

for any triangle T_0 contained in T. Our aim is to show that $|s(T)|$ is smaller than any positive number so that it would follow that $s(T) = 0$. We divide the region T into four triangles by joining the midpoints of the three sides and label them as $T^{(1)}, \ldots, T^{(4)}$. Observe that $s(T) = \sum_{j=1}^{4} s(T^{(j)})$,

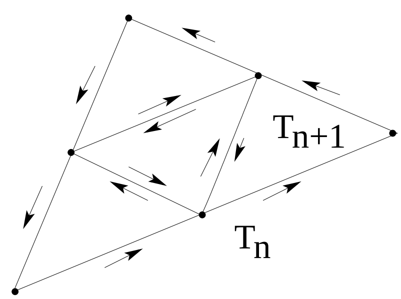


Fig. 20

since the integrals taken over the segments of $\partial T^{(j)}$ which are in the interior of T cancel in pairs as each of them occurs once in each of the two directions, whereas the integrals over the segments which make up the boundary of T occur only once on either side and with the same orientation. [This argument should be noted and learnt properly for future use also.]

Therefore, we have,

$$\left|s(T^{(j)})\right| \;\geq\; \frac{|s(T)|}{4} \tag{4.26}$$

for at least one of $j = 1, \ldots, 4$. We select the first one that satisfies this property and denote the corresponding sub-triangle by T_1.

What we have done is to set up a chain-process: we can replace T by T_1 and repeat this process and then denote the sub-triangle obtained by this process by T_2. Inductively, we would obtain a nested sequence of triangles

$$T := T_0 \supset T_1 \supset \cdots \supset T_n \supset \cdots \tag{4.27}$$

such that

$$|s(T_n)| \geq \frac{|s(T_{n-1})|}{4} \geq \cdots \geq \frac{|s(T)|}{4^n}. \tag{4.28}$$

Recall that the diameter of a triangle is equal to the length of the longest side. Therefore it follows that if d_n, L_n denote the diameter and perimeter of T_n respectively, then we have

$$d_n \leq \frac{d_0}{2^n}, \quad L_n \leq \frac{L_0}{2^n}.$$

Now suppose z_n is one of the three vertices of T_n. Then $\{z_n\}$ is a Cauchy's sequence and hence has a limit say w. Clearly, this limit point will be in all the triangles T_n.

Now given $\epsilon > 0$, choose $\delta > 0$ such that $B_\delta(w) \subset \Omega$ and such that

$$|z - w| < \delta \implies |f(z) - f(w) - (z - w)f'(w)| < \epsilon |z - w|.$$

(This is where the complex differentiability of f has been used in the form of increment theorem.) Choose n sufficiently large so that $d_n < \delta$. Since $w \in T_n$, it follows that $T_n \subset B_\delta(w)$. Now as we have already seen

$$\int_\omega z^m \, dz = 0$$

for all non negative integers m and for all closed contours ω since z^m has a primitive defined throughout $\mathbb{C}$. In particular $\int_{\partial T_n} [f(w) + (z - w)f'(w)] \, dz = 0$ for all n. Therefore

$$s(T_n) = \int_{\partial T_n} [f(z) - f(w) - (z - w)f'(w)] \, dz$$

and hence

$$|s(T_n)| \leq \epsilon \int_{\partial T_n} |(z - w) \, dz| \leq \epsilon \, d_n \int_{\partial T_n} |dz| = \epsilon d_n L_n = \frac{\epsilon d_0 L_0}{4^n}.$$

Hence

$$|s(T)| \leq 4^n |s(T_n)| \leq \epsilon d_0 L_0.$$

Since $d_0 L_0$ is a fixed positive number and $\epsilon > 0$ is arbitrary, it follows that $|s(T)|$ is smaller that any positive real number, and hence $s(T) = 0$. ♠

Remark 4.3.1 We now introduce a simple topological notion which will help us enormously in a technical way in generalizing theorem 4.3.1.

Definition 4.3.1 Let X be a topological space, $A \subset X$. A point $a \in A$ is called an isolated point of A if there exists an open set U such that $U \cap A = \{a\}$. A is called a *discrete subset* or an *isolated subset* of X if A is closed in X and each $a \in A$ is an isolated point of A.

Remark 4.3.2

1. Any finite subset $A \subset X = \mathbb{R}^n$ is a discrete subset.

2. The set of points $\{1/n \ : \ n \geq 1\}$ is a discrete subset of the open interval $(0,1)$ but not a discrete subset of $[0,1]$, just because it is not closed in the latter.

3. A discrete subset cannot have a limit point in X. In particular, a discrete subset of a closed and bounded set X is necessarily finite. (See theorem 1.5.3.)

4. Any subset of a discrete subset is discrete.

Remark 4.3.3 Given a complex differentiable function, one would like to know what is the largest subset of $\mathbb{C}$ on which it is defined. This question itself is somewhat vague in the sense that strictly speaking when a function is given it comes with its domain of definition. But then one can artificially restrict the domain of a function to get a different function. We would like to treat two such functions as one and the same. There is a deep reason to do so especially in case of complex differentiable functions. This will become clear a little later. At present to handle this question, we shall make a definition.

Definition 4.3.2 Let A be a domain in $\mathbb{C}$, $f : A \to \mathbb{C}$ be a complex differentiable function. Then all points in $\mathbb{C} \setminus A$ are called singularities of f. The isolated points of A are called isolated singularities. If there exists another complex differentiable function $g : B \to \mathbb{C}$ such that $A \subset B$ and $g(a) = f(a)$ for all $a \in A$, then we call points of $B \setminus A$ as removable singularities of f.

Remark 4.3.4 In other words, a point a is an isolated singularity of a complex differentiable function if we know that the function is defined and complex differentiable at all points of a disc around a except perhaps at a. It can happen that f is defined and complex differentiable even at a also but **we** do not know about it. By exploiting certain mild properties of f around such points we shall indeed recover this knowledge.

Theorem 4.3.2 *Let f be a continuous function in a region Ω and complex differentiable in $\Omega \setminus A$ where A is a discrete subset of Ω. Let T be a triangle completely contained in Ω. Then*

$$\int_{\partial T} f(z)\, dz = 0.$$

Proof: Observe that T being closed and bounded and A being discrete subset of Ω, it follows that $T \cap A$ is a finite set. Cut T into finitely many smaller triangles T_k so that all the points of $A \cap T = \{\xi_1, \ldots, \xi_k\}$ lie on the edges of one of these smaller triangles. (There are so many ways to do this. For instance, at any stage, pick any point ξ_j which happens to be an interior point of a triangle T_k and cut the triangle T_k into three triangles by joining ξ_j to the vertices of T_k. Now the number of ξ_j which are in the interior has gone down. You can now repeat this process finitely many times. See Fig. 21.) Then, as seen before,

$$s(T) = \sum_k s(T_k)$$

and hence it is enough to prove that each $s(T_k) = 0$.

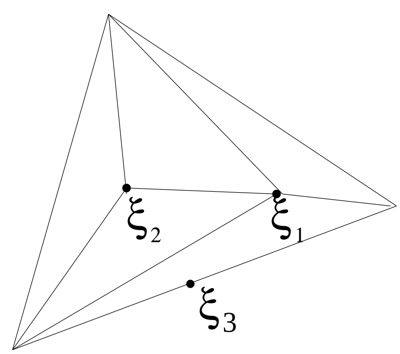


Fig. 21

Hence we could have as well assumed that all the point ξ_j belong to the boundary of T itself.

Say, $T = \Delta(a, b, c,)$. Choose three sequences a_n, b_n, c_n in the interior of T converging respectively to a, b, c. Now each triangle $T_n := \Delta(a_n, b_n, c_n)$ is contained completely in the interior of T. By the previous theorem,

$$\int_{\partial T_n} f(z) dz = 0, \quad \forall\ n.$$

Recall that the line segments $[a_n, b_n]$ are parameterised by

$$\alpha_n : t \mapsto (1-t)a_n + t b_n.$$

Check that the sequence of functions $\{\alpha_n\}$ converges uniformly to

$$\alpha : t \mapsto (1-t)a + tb$$

which is the standard parameterization of the segment $[a,b]$. Therefore

$$\int_{[a_n,b_n]} f(z)dz = \int_0^1 f(\alpha_n(t))\alpha_n'(t)dt = \int_0^1 f(\alpha_n(t))(b_n - a_n)dt.$$

Upon taking the limit under the integral sign, we have,

$$\lim_{n\to\infty} \int_{[a_n,b_n]} f(z)dz = \int_0^1 f(\alpha(t))(b-a)dt = \int_0^1 f(\alpha(t))\alpha'(t)dt = \int_{[a,b]} f(z)dz.$$

Therefore

$$0 = \lim_{n\to\infty} \int_{\partial T_n} f(z)dz = \int_{\partial T} f(z)dz.$$

This completes the proof. ♠

Remark 4.3.5 Classically, and in most of the literature, Cauchy-Goursat theorem 4.3.1 and its extension 4.3.2 are stated and proved for rectangles. This is an easy consequence of theorem 4.3.1 or 4.3.2 accordingly. Also, the latter one is stated with a seemingly weaker hypothesis that at finitely many points ξ_j of the rectangle, we have

$$\lim_{z\longrightarrow\xi_j} (z-\xi_j)f(z) = 0$$

instead of continuity of f. Our approach offers a lot of simplicity of the exposition. Also, we shall be able to recover this seemingly stronger form of Cauchy-Goursat theorem later, without additional efforts.

Theorem 4.3.3 Cauchy's Theorem on a Convex Region: *Let U be a convex region and A be a discrete subset of U. Let f be a continuous function on U and complex differentiable on $U \setminus A$. Then for any closed contour ω in U we have,*

$$\int_\omega f(z)\, dz = 0.$$

Proof: Apply corollary 4.2.1, with $p = f$ and $q = \imath f$. Given any triangle T, since T is closed and bounded, only finitely many points of A can be in T. Therefore, by Theorem 4.3.2, we have

$$\int_{\partial T} pdx + qdy = \int_{\partial T} f(z)dz = 0.$$

Therefore $\int_\omega f(z)dz = 0$ for all closed contours ω in U. ♠

Exercise 4.3

1. Evaluate the integrals around the unit circle taken counterclockwise by using Cauchy's theorem, whenever it is valid. In each case, give reasons why you can or cannot use Cauchy's theorem.
   (a) $|z|$; (b) $\mathrm{Ln}\,(z+3)$; (c) $\dfrac{1}{|z|^5}$; (d) e^{-z^2}; (e) $\tanh z$; (f) $\bar{z}$; (g) $\dfrac{1}{z^3}$.

2. Evaluate
   (a) $\displaystyle\int_C \frac{z^2-z+2}{z^3-2z^2}dz$, where C is the boundary of the rectangle with vertices $3\pm i, -1\pm i$ traversed clockwise.
   (b) $\displaystyle\int_C \frac{\sin z}{z+3i}dz$, $C : |z-2+3i| = 1$ (counterclockwise)

3. C is the unit circle traversed counterclockwise. Integrate over C,
   (a) $\dfrac{e^z-1}{z}$ (b) $\dfrac{z^3}{2z-i}$ (c) $\dfrac{\cos z}{z-\pi}$ (d) $\dfrac{\sin z}{2z}$.

4.4 * Cauchy's Theorem via Green's Theorem

Cauchy's definition of a holomorphic function (he called it 'synectic function'), as given in definition 3.3.1, included the stronger hypothesis of continuity of the first order partial derivatives along with complex differentiability. He used variational principle to arrive at a proof of his theorem. In this section we outline a proof which uses Green's theorem. The only serious snag in this approach is that we will not see a proof of the fact :*complex differentiability in an open set implies holomorphicity.* The other and not so serious snag is that this approach implicitly depends on Jordan curve theorem(JCT) which we are not going to prove. However, for all practical instances where we have to appeal to JCT, it has been possible to give an ad hoc proof of the necessary conclusion of JCT and so the snag is not so serious.

Definition 4.4.1 An open set Ω in $\mathbb{C}$ is called (geometrically) simply connected if for every simple closed contour ω in Ω, the inside region of ω is contained in Ω.

Remark 4.4.1 Recall that by Jordan curve theorem (JCT) (1.6.3), ω separates $\mathbb{C}$ into two components one unbounded and another bounded. The bounded component is called the inside of ω. Thus the above definition is equivalent to say that points outside Ω are not 'enclosed' by any simple closed contour in Ω. Obviously, the entire plane $\mathbb{C}$ is simply connected, since there is no outside point at all. Also if A is a non empty bounded set, then it follows that $\mathbb{C} \setminus A$ is not simply connected, for we can merely take ω to be the circle $|z| = M$ where M is such that $|a| < M$ for all $a \in A$. Then the inside of ω is not contained in $\mathbb{C} \setminus A$. In particular, it follows from JCT that the outside of a simple closed contour ω in $\mathbb{C}$ is not simply connected. On the other hand, it is not so hard to see that the inside of ω is simply connected. This we leave it to you as an exercise. We also leave it to you as an exercise that any convex open set in $\mathbb{C}$ is connected and simply connected.

Theorem 4.4.1 Cauchy's Theorem I-version *Let Ω be a simply connected domain in $\mathbb{C}$ and f be a holomorphic function on it. Then for any simple closed contour γ in Ω we have*

$$\int_\gamma f(z)dz = 0.$$

Proof: Let R be the domain bounded by γ. Since Ω is simply connected, it follows that $R \subset \Omega$. Therefore, if $f = u + \imath v$, then u, v have continuous partial derivatives at all points of $R \cup \gamma$. Moreover,

$$f(z)dz = (udx - vdy) + i(vdx + udy).$$

Therefore by Green's [2] theorem, we have,

$$\begin{aligned} \int_\gamma f(z)dz &= \int_{\partial R} f(z)dz \\ &= -\iint_R (u_y + v_x)dxdy + i\iint_R (u_x - v_y)dxdy \end{aligned}$$

Since u, v satisfy C-R equations throughout Ω, the integrands in both the double integrals above vanish identically. ♠

[2] George Green(1793-1841), an English Mathematician without any formal training, was a baker to begin with and became a fellow of Caius College Cambridge. He worked on potential theory of electricity and magnetism, waves and elasticity. We use the following version of Green's Theorem (see [K] for a proof): *Let $u(x, y), v(x, y)$ be continuous functions with continuous partial derivatives on an open set containing a closed and bounded region R with the boundary ∂R consisting of finitely many piecewise smooth curves each oriented in such a way that the region R always lies on the left of the curve. Then*

$$\iint_R (u_x - v_y)dxdy = \int_{\partial R} (udy + vdx).$$

Remark 4.4.2 Observe that we have used continuity of u_x, u_y, v_x, v_y above in employing Green's theorem.

Using Green's theorem for multi-connected domains, allowing the boundary to be finite union of simple closed contours, and arguing as before, we obtain the following:

Theorem 4.4.2 Cauchy's Theorem II-version *Let R be a domain in $\mathbb{C}$ bounded by finitely many simple closed, oriented contours ∂R. Suppose f is holomorphic on an open set Ω containing $R \cup \partial R$. Then*

$$\int_{\partial R} f(z)dz = 0.$$

Next we slacken the condition on the function f.

Theorem 4.4.3 Cauchy's Theorem III-version *Let Ω be a simply connected domain and $A \subset \Omega$ be a finite subset. Let $f : \Omega \to \mathbb{C}$ be a continuous function such that $f : \Omega \setminus A \to \mathbb{C}$ is holomorphic. Then for any closed contour γ in Ω, we have*

$$\int_{\gamma} f(z)dz = 0. \tag{4.29}$$

Proof: By expressing a given closed contour into a sum of finitely many simple ones, observe that it is enough to prove (4.29) for a simple closed contour γ.

We shall first prove this for the case when γ does not pass through any points of A.

Let R be the domain enclosed by γ. Then $R \subset \Omega$. This is precisely where simple connectivity of Ω is used. Let $A \cap R = \{a_1, \ldots, a_k\}$. Given $\epsilon > 0$, we must show that $\left| \int_{\gamma} f(z)dz \right| \leq \epsilon$. Let M be an upper bound for $|f(z)|$ on R. Choose $0 < r < \frac{\epsilon}{2\pi k M}$ such that $B_r(a) \cap \gamma = \emptyset$. Put $S = R \setminus \cup_{j=1}^{k} B_r(a_j)$. By the II-version of Cauchy's theorem applied to f on the domain S, we obtain

$$\int_{\partial S} f(z)dz = 0.$$

Let C_j be the oriented boundary of $B_r(a_j)$. Since $\partial S = \gamma \cup (C_1)^{-1} \cup (C_2)^{-1} \cup \cdots \cup (C_k)^{-1}$, we get

$$\int_{\gamma} f(z)dz = \sum_{j=1}^{k} \int_{C_j} f(z)dz. \tag{4.30}$$

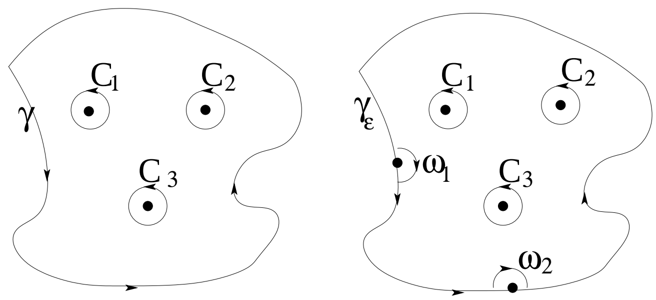


Fig. 22

Now by M-L inequality, it follows that

$$\begin{aligned}\left|\int_{\gamma} f(z)dz\right| &= \left|\sum_{j=1}^{k}\int_{C_j} f(z)dz\right| \\ &\leq \sum_{j=1}^{k} ML(C_j) = 2\pi rkM \leq \epsilon.\end{aligned}$$

Next, we can generalize this to the case when γ is a simple closed curve which may pass through $\{b_1, \ldots, b_p\} \subset A$. Choose $\epsilon > 0$ so that the circles $|z - b_j| = \epsilon$ are contained in Ω. While tracing the curve γ as we reach a point on any of these circles, use some arcs ω_j of the circle to go around the point and avoid tracing the part γ_j of the curve γ lying inside the circle $|z - b_j| = \epsilon$. If γ_ϵ is the simple closed curve so obtained then

$$\int_{\gamma} f(z)dz = \int_{\gamma_\epsilon} f(z)dz - \sum_j \int_{\omega_j} f(z)dz + \sum_j \int_{\gamma_j} f(z)dz.$$

The first integral on the RHS vanishes because it is a simple closed curve in Ω avoiding the points in A. Using M-L inequality, we can bound each of the integral in the two summations by a number $2\pi M\epsilon$. Since ϵ can be chosen arbitrarily small, it follows that $\displaystyle\int_{\gamma} f(z)dz = 0$. ♠

Exercise 4.4

1. Solve all the exercises in 4.3.

2. Evaluate $\displaystyle\int_B f(z)dz$ where $f(z)$ is given by
(a) $\dfrac{z+2}{\sin\frac{z}{2}}$; (b) $\dfrac{z}{1-e^z}$,
where B is the boundary of the domain between $|z| = 4$ and the square with sides along $x = \pm 1, y = \pm 1$, oriented in such a way that the domain always lies to its left.

3. If ω is a simple closed contour in $\mathbb{C}$, show that the inside region of ω is simply connected.

4.5 Cauchy's Integral Formulae

Recall the result in (4.20). In contrast, we now have

$$\int_{|z-a|=r} \frac{dz}{z-w} = 0$$

for points w outside the disc $|z-a| \leq r$.

Proving a vanishing theorem of this type is one thing. Using this to obtain an integral formula is another thing. This is one place where Cauchy's ingenuity is beyond any doubt. We begin with:

Proposition 4.5.1 Let A be a discrete subset of a convex region Ω. Let f be a continuous function on Ω and holomorphic on $\Omega' := \Omega \setminus A$. Let ω be a closed contour in Ω. Then for any point $z_0 \in \Omega' \setminus Im(\omega)$,

$$\boxed{\int_\omega \frac{f(z)\,dz}{z-z_0} = f(z_0) \int_\omega \frac{dz}{z-z_0}.} \tag{4.31}$$

Proof: Consider the following function on $\Omega \setminus \{z_0\}$,

$$F(z) = \begin{cases} \dfrac{f(z)-f(z_0)}{z-z_0}, & z \neq z_0 \\ f'(z_0), & z = z_0. \end{cases}$$

This is holomorphic on $\Omega' \setminus \{z_0\}$ and continuous in Ω. Hence, by theorem 4.3.3

$$\int_\omega F(z)\,dz = 0$$

which yields (4.31). ♠

Theorem 4.5.1 Cauchy's integral formula on a disc: *Let f be a complex differentiable function on an open set containing the closure of a disc D. Then*

$$\boxed{f(z) = \frac{1}{2\pi\imath} \int_{\partial D} \frac{f(\xi)}{\xi - z} d\xi, \ \ z \in D.} \tag{4.32}$$

Proof: In the Proposition 4.5.1, we take the contour of integration to be positively oriented boundary of the disc D. By (4.20) we then have

$$\int_{\partial D} \frac{d\xi}{\xi - z} = 2\pi\imath.$$

(4.32) follows. ♠

Example 4.5.1 Let us find the value of

$$\int_{|z|=1} \frac{e^{az}}{z}\, dz.$$

Observe that e^{az} is complex differentiable on the entire plane $\mathbb{C}$. Therefore, take $f(z) = e^{az}$ and the disc D to be the unit disc in (4.32). It follows that the integral is equal to $2\pi\imath f(0) = 2\pi\imath$.

The advantage of representing a complex differentiable function as an integral is tremendous. With this result, we have arrived at one of the peaks. We can now take a look at various directions. The integral formula helps in arriving at several important results. In the next three sections we give two such samples. More of it will come in the next chapter.

Let us just play around with this great formula (4.32). For instance, let us just differentiate under the integral sign repeatedly.

Theorem 4.5.2 Cauchy's Integral formula for Derivatives: *Let f be complex differentiable in a region Ω. Then f has complex derivatives of all order in Ω. Moreover, if D is a disc whose closure is contained in Ω and $z \in \operatorname{int} D$ then for all integers $n \geq 0$, we have,*

$$\boxed{f^{(n)}(z) = \frac{n!}{2\pi\imath} \int_{\partial D} \frac{f(\xi)\, d\xi}{(\xi - z)^{n+1}}} \tag{4.33}$$

where ∂D denotes the boundary circle traced in the counter clockwise sense.

Proof: It is enough to prove the formula (4.33), which we shall do by induction on n. For $n = 0$, this is the same as (4.32). So assume (4.33) holds for $n - 1$ and for all points inside D. Differentiating (4.33) under the integral sign gives (4.33) for n.

In particular, it follows that f has derivatives of all order inside the disc D. Since for every point z in Ω, we can find a closed disc $D \subset \Omega$ such that $z \in \operatorname{int} D$, the same holds for all point of Ω. ♠

Remark 4.5.1 Thus, it follows that a complex differentiable function in a domain is differentiable any number of times. In particular, the difference in the classical definition and our definition of holomorphicity vanishes. We end this section with a handy criterion for holomorphicity for continuous functions.

Theorem 4.5.3 Morera's[3] Theorem : *If f is a continuous function on a region Ω such that for all triangles $T \subset \Omega$, $\displaystyle\int_{\partial T} f\,dz = 0$. Then f is complex differentiable in Ω.*

Proof: For the purpose of proving complex differentiability of f, we can restrict the domain to a disc and assume that Ω is a disc. By corollary 4.2.1, it follows that the hypothesis for corollary 4.2.2 is satisfied. Therefore f has a primitive on Ω, i.e., there exists a complex differentiable function F on Ω such that $F'(z) = f(z),\ \forall\ z \in \Omega$. The function F, being complex differentiable in Ω has derivatives of all order, by the above theorem. Therefore $f = F'$ also has derivatives of all order. ♠

Exercise 4.5

1. Integrate $\dfrac{1}{z^4 - 1}$ over (a) $|z + 1| = 1$, (b) $|z - i| = 1$, each curve being taken counterclockwise. [Hint: Resolve into partial fractions.]

2. Let C be the circle $|z| = 3$ traced in the counterclockwise sense. For any z with $|z| \neq 3$, let $g(z) = \displaystyle\int_C \frac{2w^2 - w - 2}{w - z} dw$. Prove that $g(2) = 8\pi i$. Find $g(4)$.

3. Prove a result similar to theorem 4.3.2 except that A is a line segment rather than a discrete subset. [Hint: Follow similar line of argument as in the theorem, viz., cut down the triangle to avoid the line segment from being part of the interior of the triangle.]

4. Let Ω be a convex domain and $f : \Omega \to \mathbb{C}$ be a continuous function. If $A \subset \Omega$ is a line segment such that f is complex differentiable on $\Omega \setminus A$, then show that f is complex differentiable on Ω.

5. In the above exercise can one take A to be an arc of a circle? How far can you generalize this?

[3]Giacinto Morera(1856-1909) was an Italian mathematician. He proved this theorem in 1886.

6. **Schwarz's Reflection Principle:** Let Ω be a convex domain which is closed under conjugation, i.e., such that $z \in \Omega$ iff $\bar{z} \in \Omega$. Let

$$\Omega^+ = \{z \in \Omega \ : \ \Re(z) > 0\}; \quad \Omega^- = \{z \in \Omega \ : \ \Re(z) < 0\}$$

and $A = \Omega \cap \mathbb{R}$. Let $f : \Omega^+ \cup A \to \mathbb{C}$ be a continuous function such that f is complex differentiable in Ω^+ and $f(A) \subset \mathbb{R}$. Define $g : \Omega \to \mathbb{C}$ by

$$g(z) = \begin{cases} f(z), & z \in \Omega^+ \cup A; \\ \overline{f(\bar{z})}, & z \in \Omega^-. \end{cases}$$

Show that g is complex differentiable in Ω.

7. Let Ω be a region which is closed under conjugation. Show that every complex differentiable function f on Ω can be expressed as a sum $f_1 + \imath f_2$, where f_j are complex differentiable and map the real axis into itself.

8. Remove the hypothesis that $f(A) \subset \mathbb{R}$ in Ex. 6.

   [We shall prove a stronger version of the result in Ex. 6-8, in section 4.9 for harmonic functions.]

9. State a result similar Ex.6 for the unit circle instead of the real line and prove it.

4.6 Analyticity of Complex Differentiable Functions

Remark 4.6.1 *Behold! We have shown that a function which is once complex differentiable is differentiable any number of times.* Certainly this is something that we never bargained for while launching the theory of complex differentiation. There is more to come. It is time for us to reap the harvest. By a simple use of geometric expansion, we shall now prove the result that we have promised long back, which will allow us to use the phrases complex differentiability, holomorphicity and analyticity to mean the same class of functions.

Theorem 4.6.1 Analyticity of Complex Differentiable Functions *Let $f : \Omega \to \mathbb{C}$ be a complex differentiable function, D be the open disc with center a and radius r whose closure is contained in Ω. Then f has the power series expansion*

$$f(z) = \sum_0^\infty \frac{f^{(n)}(a)}{n!}(z-a)^n \tag{4.34}$$

valid for all points inside D.

Proof: For points ξ, z such that $|\xi - a| = r$, and $|z - a| < r$, we have

$$\frac{1}{\xi - z} = \frac{1}{(\xi - a) - (z - a)} = \frac{1}{\xi - a}\sum_{n=0}^{\infty}\left(\frac{z-a}{\xi - a}\right)^n$$

which is uniformly convergent in the closed disc $|z - a| \leq r'$ for $r' < r$. We now take Cauchy integral formula (4.32), substitute the above series expansion for $\frac{1}{\xi - z}$ on the RHS, use the uniform convergence to interchange the order of integration and summation to obtain a power series in $(z - a)$, and identify the coefficients of $(z - a)^n$ using Cauchy's formula for derivatives (4.33). Thus, we have

$$\begin{aligned} f(z) &= \frac{1}{2\pi\imath}\int_{|\xi - a| = r}\frac{f(\xi)}{\xi - z}d\xi \\ &= \frac{1}{2\pi\imath}\int_{|\xi - a| = r}\left(\sum_{n=0}^{\infty}\frac{f(\xi)}{(\xi - a)^{n+1}}(z-a)^n\right)d\xi \\ &= \sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n. \end{aligned}$$

Since this is true for all $|z - a| \leq r' < r$, this true for all $z \in D$. ♠

Remark 4.6.2 The above series (4.34) is called the Taylor's series of f. If $a = 0$, the same goes under the name Maclaurin's series. Clearly, for any point $a \in \Omega$, the above power series expansion is valid for the biggest disc that is contained in Ω and having center at a. It may well happen that the actual radius of convergence of this series is even bigger than the radius of this biggest disc. The remainder after n terms is obviously an analytic function. A slight modification of the above arguments yields an integral formula for the remainder. This result goes under the name Taylor's Formula:

Theorem 4.6.2 Taylor's Formula: *Let $f : \Omega \to \mathbb{C}$ be a complex differentiable function, and $a \in \Omega$. Let $r > 0$ be such that $B_r(a) \subset \Omega$. Then, for all $z \in B_r(a)$,*

$$\boxed{f(z) = f(a) + f'(a)(z-a) + \cdots + \frac{f^{(n)}(a)}{n!}(z-a)^n + \phi(z)(z-a)^{n+1},} \tag{4.35}$$

where,

$$\boxed{\phi(z) = \frac{1}{2\pi\imath}\int_{|\xi - a| = r}\frac{f(\xi)}{(\xi - a)^{n+1}(\xi - z)}d\xi.} \tag{4.36}$$

Proof: We begin with the algebraic identity

$$1 + t + t^2 + \cdots + t^n = \frac{1 - t^{n+1}}{1 - t}$$

and substitute $t = \dfrac{z - a}{\xi - a}$ to obtain

$$1 + \frac{z-a}{\xi - a} + \cdots + \left(\frac{z-a}{\xi-a}\right)^n = \frac{\xi - a}{\xi - z} - \frac{(z-a)^{n+1}}{(\xi - z)(\xi - a)^n}.$$

Upon multiplying by $\dfrac{f(\xi)}{2\pi\imath(\xi - a)}$ throughout, integrating on $|\xi - a| = r$ and using the Cauchy formulae for $f^{(k)}, k = 0, 1, \ldots, n$, we obtain the desired result. ♠

Example 4.6.1 Newton's Binomial Series: Consider a well defined branch $f(z)$ of $(1+z)^\alpha$ in $\mathbb{D}$, where α is some real number. Of course, if α is a non negative integer, this function is univalent and hence there is no problem. So, here we consider the case when α is not a non negative integer. Then it is seen that the successive coefficients of the Taylor's series are given by

$$\boxed{(1+z)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n = \sum_{0}^{\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} z^n.} \tag{4.37}$$

Since the function f is complex differentiable in $\mathbb{D}$, it follows from the above theorem that the Taylor series has radius of convergence at least 1. On the other hand, if the radius of convergence were bigger than 1, then it would mean that all the derivatives of $(1+z)^\alpha$ are bounded at $z = -1$. This is easily seen to be false by taking n^{th}-derivative for $n > \alpha$. Therefore, we conclude that the radius of convergence is 1. As a corollary, using Cauchy-Hadamard formula for the radius of convergence, we obtain that

$$\limsup_n \left|\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}\right|^{1/n} = 1.$$

We have included this argument just to illustrate a two-way usage of this theory. If you can compute the radius of convergence, then you know the disc on which the function is analytic, whereas, if you already know the holomorphicity of the function, then you know something about the radius of convergence from which you can find the limit of the sequence of n^{th} roots of the coefficients, if it exists.

Let us find the power series representation for the inverse of *tan* function, viz., a well chosen branch of $f(z) = \arctan z$. We observe that,

$$f'(z) = \frac{1}{1+z^2} = 1 - z^2 + z^4 - \cdots + \cdots ,$$

and hence by integrating term by term, we obtain,

$$\boxed{\arctan z = z - \frac{z^3}{3} + \frac{z^5}{5} - \cdots + \cdots .} \tag{4.38}$$

The radius of convergence is easily seen to be 1.

Remark 4.6.3 Having proved that $\mathbb{C}$-differentiable functions are analytic and hence holomorphic, we can now use results of section 4.4 directly. In particular, the II and III version of Cauchy's theorem are now available to us via Green's theorem. However, in the next chapter we shall prove sweeping generalizations of these results.

Exercise 4.6

1. Let $f(z) = \sum a_n z^n$ be a convergent power series. Suppose f is an odd (respectively an even) function in the disc. Show that $a_{2n} = 0$ (respectively, $a_{2n-1} = 0$) for all $n \geq 1$.

2. Find power series representations for the following functions choosing the centers and the branches appropriately whenever applicable and find the radius of convergence of the power series that you have obtained. Also compute the coefficients up to say n= 5 at least.
   (i) $(1 - z^2)^{1/2}$, (ii) $(1 - z^2)^{-1/2}$ (iii) $\arcsin z$.

3. Consider the Maclaurin's series expansion

$$\boxed{g(z) = \tfrac{z}{e^z - 1} = \sum_0^\infty \tfrac{B_n}{n!} z^n.} \tag{4.39}$$

   (The complex numbers B_n are called the *Bernoulli*[4] *numbers*. These are very important in the study of analytic, number theoretic and algebro-geometric problems.) Using the fact that $g(z) + z/2$ is an even function conclude that $B_1 = -1/2$ and $B_{2n+1} = 0$, $n \geq 1$. Also by comparing the coefficients of the identity

[4]Jacob Bernoulli(1665-1705), a Swiss mathematician found these numbers while computing the sums of powers of integers. He is the senior most amongst four famous Bernoullies.)

$$1 = \left(\frac{e^z - 1}{z}\right)\left(\frac{z}{e^z - 1}\right) = \left(\sum_1^\infty \frac{z^{n-1}}{n!}\right)\left(\sum_0^\infty \frac{B_n}{n!} z^n\right),$$

prove that

$$B_0 + \binom{n}{1} B_1 + \binom{n}{2} B_2 + \cdots + \binom{n}{n-1} B_{n-1} = 0.$$

Compute the Bernoulli numbers say, up to B_{16}. What is the radius of convergence of the series (4.39)? From this, obtain the value of $\limsup_n \sqrt[n]{\frac{B_n}{n!}}$.

4. Show that the Maclaurin's series for $z \cot z$ and $\tan z$ are given by

$$\boxed{z \cot z = 1 + \sum_1^\infty (-1)^n \frac{2^{2n} B_{2n}}{(2n)!} z^{2n}.} \tag{4.40}$$

$$\boxed{\tan z = \sum_1^\infty (-1)^{n-1} \frac{2^{2n}(2^{2n} - 1) B_{2n}}{(2n)!} z^{2n-1}.} \tag{4.41}$$

Determine the radius of convergence in each case.

[Hint: First express $z \cot z$ in terms of $g(z)$ as given in Ex.3 above and use the expansion for $g(z)$. For $\tan z$ use the identity $\tan z = \cot z - 2 \cot 2z$.]

5. Obtain the Maclaurin series for $\ln\left(\frac{\sin z}{z}\right)$. Find the radius of convergence.

6. Let $P(t) = \sum_n a_n t^n$ be a power series such that $P(z - 1/2) = \frac{1}{1-z}$ in some neighborhood of $1/2$. What is the radius of convergence of P? What is the value of a_5?

7. Show that $\lim_{y \longrightarrow \infty} \imath y \left[\pi \imath + \ln\left(\frac{1/2 - \imath y}{1/2 + \imath y}\right)\right] = -1$. [Hint: Pull out $\imath y$ and use logarithmic expansion.]

4.7 A Global Implication: Liouville

Let us now have a sample of the kind of implications CIF has on global behavior of complex differentiable functions.

Theorem 4.7.1 Cauchy's Estimate : *Let f be complex differentiable in an open set containing the closure of the disc $B_r(z)$ and let $M_r = Sup\{|f(\xi)| \ : \ |\xi - z| = r\}$. Then for all $n \geq 1$ we have,*

$$|f^{(n)}(z)| \leq \frac{n! M_r}{r^n}. \tag{4.42}$$

Proof: Take C to be the circle of radius r around z. Then we have,

$$|f^{(n)}(z)| = \left| \frac{n!}{2\pi\imath} \int_C \frac{f(\xi)\, d\xi}{(\xi - z)^{n+1}} \right| \leq \frac{n! M_r}{2\pi r^{n+1}} \int_C |d\xi| = \frac{n! M_r}{r^n}.$$

This proves the theorem. ♠

Definition 4.7.1 A function that is complex differentiable on the entire plane $\mathbb{C}$ is called an *entire function.*

Theorem 4.7.2 Liouville's[5] Theorem : *A bounded entire function is a constant.*

Proof: Putting $n = 1$ in the Cauchy's estimate, we obtain that $|f'(z)| \leq M_r/r$. Since f is bounded, let M be such that $M_r \leq M$ for all r. Now take the limit as $r \longrightarrow \infty$ to conclude that $f'(z) = 0$. Since z was arbitrary point, this implies f is a constant. ♠

As a consequence of Liouville's theorem, we shall now prove the FTA, thereby fulfilling an old promise. (Compare this with the proof given in section 1.7.)

Theorem 4.7.3 The Fundamental Theorem of Algebra: *Let $p(z) = a_n z^n + \cdots + a_1 z + a_0$, $a_j \in \mathbb{C}, a_n \neq 0$ be a polynomial function in one variable of degree $n \geq 1$ over the complex numbers. Then the equation $p(z) = 0$ has at least one solution in $\mathbb{C}$.*

Proof: Assume that $p(z)$ is never zero. Then as seen before, it follows that $f(z) = 1/p(z)$ is differentiable everywhere, i.e., $f(z)$ is an entire function. We shall show that $f(z)$ is bounded and then from Liouville's theorem it follows that f is a constant and hence p is a constant. But it is easily verified that any polynomial function of positive degree is not a constant. This contradiction will prove the theorem. So to show that f is bounded, recall Ex. 11 of section 1.1 (or prove it afresh) that $|p(z)|$ tends to infinity as $|z|$ tends to infinity, hence we can find large r such that $|z| > r \Longrightarrow |f(z)| < 1$. On the other hand, the continuity of f gives you a bound for $f(z)$ inside the disc $|z| \leq r$. This shows that f is bounded as claimed. ♠

[5]Joseph Liouville(1809-1882) was a French mathematician. He gave a proof of the above theorem while lecturing on his work on doubly periodic function in 1847. A German mathematician C. W. Borchardt who heard this lecture published this result attributing it to Liouville. However, Cauchy had already derived it in 1844, using his calculus of residues.

Exercise 4.7

1. Given a complex differentiable function f such that $|f(z)| \leq 1, \ \forall \ |z| \leq 1$, find a variable upper bound for $|f^{(n)}(z)|$ in the discs $|z| < 1$. Conclude that $|f^{(n)}(0)| \leq n!$.

2. Let f be a complex differentiable function on the unit disc such that $|f(z)| < (1 - |z|)^{-1}$ for $|z| < 1$. Use Cauchy's estimate to show that $|f^{(n)}(0)| < (n+1)!e$.

3. Show that if f is an entire function such that $|f(z)| \leq k|z^n|, |z| > M$ for some constant k, M and some positive integer n, then f is a polynomial function. Find an estimate for the coefficient of the top degree term.

4.8 Mean Value and Maximum Modulus

In this section we give a sample of mixture of local and global behavior.

Let us begin with:

Corollary 4.8.1 Gauss's Mean Value Theorem Let f be a complex differentiable function on a disc $B_R(z_0)$. Then for $0 < r < R$,

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + re^{\imath\theta}) d\theta. \tag{4.43}$$

Proof: This is just going back to the definition of the right hand side of the formula

$$f(z_0) = \frac{1}{2\pi\imath} \int_{|z-z_0|=r} \frac{f(z)}{z - z_0} dz.$$

The parameterization of the circle is $z(\theta) = z_0 + re^{\imath\theta}, 0 \leq \theta \leq 2\pi$, and therefore the rhs is given by

$$\frac{1}{2\pi\imath} \int_0^{2\pi} \frac{f(z_0 + re^{\imath\theta})}{re^{\imath\theta}} re^{\imath\theta} \imath d\theta = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + re^{\imath\theta}) d\theta.$$

This is obviously the *arithmetic mean* (continuous version!) of the function f on the circle. This is in some sense much more stronger than the Lagrange Mean value theorem of real 1-variable calculus, which only says that the mean value is attained at some point in the interval. By taking real and imaginary parts we get mean value property of harmonic functions (more about this in next section). Any kind of mean value theorems are going to be useful. Here is an illustration.

Theorem 4.8.1 Maximum Modulus Principle *Let $f : \Omega \to \mathbb{C}$ be a* **non constant** *complex differentiable function on a domain Ω. Then there does not exist any point $w \in \Omega$ such that $|f(z)| \leq |f(w)|$ for all $z \in \Omega$.*

Proof: If possible, let there be such a point. Let A be the set of all such $w \in \Omega$. By the assumption A is non empty. Say $w_0 \in A$ and $k = |f(w_0)|$. Then $A = \{w \in \Omega : |f(w)| = k\}$. Hence A is a closed set. We shall prove that A is an open set also. Then from theorem 1.6.2, it follows that A is the whole of Ω. But then $|f|$ is a constant on Ω. By example 3.1.2, f is a constant which contradicts the hypothesis.

To prove that A is open, let $a \in A$ and choose $r > 0$ such that $B_r(a) \subset \Omega$. Then for $0 < r' < r$, we have,

$$f(a) = \frac{1}{2\pi} \int_0^{2\pi} f(a + r'e^{\imath\theta})d\theta.$$

Therefore

$$k = |f(a)| \leq \frac{1}{2\pi} \int_0^{2\pi} |f(a + r'e^{\imath\theta})|d\theta.$$

Therefore

$$\int_0^{2\pi} (k - |f(a + r'e^{\imath\theta})|)d\theta \leq 0.$$

Since the integrand is a continuous non negative real function, this means it is identically zero, i.e., $k = |f(a + r'e^{\imath\theta})|$ for all θ. Since this is true for $0 < r' < r$, we have shown that $|f(z)| = k$ for all $z \in B_r(a)$ i.e., $B_r(a) \subset A$. Since this is true for all $a \in A$, we have proved that A is open. ♠

Remark 4.8.1

1. Suppose f is a complex differentiable function which never vanishes. Then $1/f$ is holomorphic and by maximum modulus principle applied to this, it follows that $|f|$ does not attain its minimum in the interior of the domain.

2. There are several equivalent versions of the maximum modulus principle. Here is one such. Suppose f is a non constant holomorphic function on a closed, connected and bounded set K of $\mathbb{C}$, then the maximum of $|f(z)|$ occurs only on the boundary of K. To see this, observe that since K is assumed to be closed and bounded, $|f(z)|$ being continuous, attains its maximum at some $z \in K$. However, $z \notin int\, K$, by the above theorem. Hence $z \in K \setminus int(K) = \partial K$.

3. Observe that if $K = \overline{\Omega}$, where Ω is an open set then it suffices to assume that f is holomorphic in Ω and continuous on $\overline{\Omega}$ to conclude as in (1).

4. On the other hand, consider a special case when $\overline{\Omega}$ is the closed disc, $|z| \leq R$. By continuity, f assumes its maximum on $|z| \leq R$. Hence it may be expected that a better estimate can be found for the modulus function at interior points of the region. Theorems to this effect are found to be quite useful. We choose to post-pone such finer study to the next chapter.

Exercise 4.8

1. Let f be a (non constant) holomorphic function a domain U. Suppose $\{z \in D : |f(z)| = k\}$ is the entire boundary of a domain $\Omega \subset U$. Show that f must vanish at some point in Ω.

2. Determine the maximum of the modulus of the functions on $[-1,-1] \times [1,1]$:
(i) e^z; (ii) $\cos z$; (iii) $z^2 + z + 1$.

3. Find the minimum of $1 + |z|^2$ on the unit disc $\mathbb{D}$ and see that it is non zero. Does this violate remark 4.8.1.1?

Remark 4.8.2 In the next section, we shall give another interesting applications of maximum modulus principle.

4.9 Harmonic Functions

Cauchy-Riemann equations tell us that the real and the imaginary parts of a complex differentiable function have some special properties. Apart from being inter-related they have the special property of possessing partial derivatives of all order, since this is the case with holomorphic functions. Similarly, the maximum modulus theorem tells us about certain distinct features of the modulus function of a holomorphic function. Such properties of real valued functions can be studied on their own and such a study can either be carried out using the knowledge of complex functions or independently. Also, an independent inquiry can lead to better understanding of the theory of complex functions themselves. The class of harmonic functions and the wider class of subharmonic functions substantiate this view with many such instances. From the application point view, few ideas surpass the notion of harmonic functions. In this section, we shall just touch upon this subject and promise to do a little more in a later chapter.

Definition 4.9.1 A real valued function $u = u(x, y)$ defined on a domain Ω in $\mathbb{C}$, is called *harmonic* with respect to the variables x, y, if it possesses continuous second order partial derivatives and satisfies the Laplace's equation:

$$\bigtriangledown^2 u := \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{4.44}$$

Remark 4.9.1

(i) Harmonic functions arise in the study of gravitational fields, electrostatic fields, steady-state heat conduction, incompressible fluid flows etc..

(ii) Technically, harmonic functions are very close to holomorphic functions. Given a holomorphic function $f = u + \imath v$, it is a straight forward consequence of Cauchy-Riemann equations and the property of a holomorphic function possessing continuous derivatives of all order, that u and v are both harmonic. In this case we call v the *harmonic conjugate* of u. Observe that, by considering the function $\imath f(x, y)$, it also follows that u is the harmonic conjugate of $-v$. We shall see the converse of this as a theorem.

(iii) There is nothing very special about considering only real valued functions for the definition of harmonic functions. We could even allow complex valued functions in the above definition. Then it follows that a complex function is harmonic iff its real and imaginary parts are harmonic. Thus, it suffices to treat only the real valued functions, in the study of harmonic functions.

(iv) From the linearity of the differential operator $\bigtriangledown^2$, it follows that the set of all harmonic functions on a domain forms a vector space. In particular all linear functions $ax + by$ are harmonic. However, it is not true that product of two harmonic functions is harmonic. For example, xy is harmonic but x^2y^2 is not.

(v) Harmonicity is quite a delicate property. If ϕ is a smooth real valued function of a real variable and u is harmonic, then, in general, $\phi \circ u$ need not be harmonic. Indeed, $\phi \circ u$ is harmonic for all harmonic u iff ϕ is linear(exercise). Likewise, if $f : \Omega_1 \longrightarrow \Omega_2$ is a smooth complex valued function of two real variables then $u \circ f$ need not be harmonic. However, under conformal mapping we have some positive result as we shall see soon.

(vi) **Polar coordinate form of Laplace:** Since $x = r\cos\theta,\ y = r\sin\theta$, the Laplace's equation takes the form

$$\boxed{\boxed{r\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{\partial^2 u}{\partial \theta^2} = 0.}} \tag{4.45}$$

Verify this. Observe that this form of Laplace's equation is not applicable at the origin, since the polar coordinate transformation is singular there. However, the polar coordinate form has many advantages of its own and is loved by physicists and engineers.

As an application, from (4.45), it follows easily that $\ln r = \ln |z|$ is well defined and harmonic throughout $\mathbb{C} \setminus \{0\}$. In fact as an easy exercise, prove that any function of r alone, i.e., independent of θ, is harmonic iff it is of the form $a \ln r + b$, where a, b are constants.

Example 4.9.1 Given a function $u : U \to \mathbb{R}$, is it the real part of a holomorphic function? As we have seen before, u must be a harmonic function. As an illustrative example of how to get holomorphic function, consider $u = x^2 - y^2$. Then $\dfrac{\partial u}{\partial x} = 2x$, $\dfrac{\partial u}{\partial y} = -2y$. Hence, $\nabla^2 u = 0$. So u is harmonic. Now, if v is a conjugate of u then v must satisfy $\dfrac{\partial v}{\partial x} = 2y$, $\dfrac{\partial v}{\partial y} = 2x$. Integrating the first equation we obtain

$$v = 2xy + \phi(y)$$

when $\phi(y)$ is purely a function of y. Differentiating this w.r.t. y and using the second equation, we see that $\phi'(y) = 0$ and hence $\phi = c$, a constant.

Now,

$$\begin{aligned} f(x + \imath y) &= x^2 - y^2 + \imath(2xy + c) \\ &= (x + \imath y)^2 + \imath c \end{aligned}$$

i.e., $f(z) = z^2 + \imath c$ is the most general holomorphic function with its real part $x^2 - y^2$.

The above discussion can be imitated to prove the following:

Theorem 4.9.1 *Let U be a convex region in $\mathbb{C}$ and $u : U \to \mathbb{R}$ be a harmonic function. Then $u = \Re(f)$ where f is a holomorphic function on U.*

We shall now give an algorithmic method of finding f, which is free from integration. To begin with, assume that the function u is a rational function of the two real variables x, y with real coefficients and is defined in a neighborhood of $(0, 0)$. We shall seek an elementary holomorphic function f such that $\Re f = u$. Let $g(x, y) := g(z) = \overline{f(z)}$. Then g is anti-holomorphic.

Now in the identity

$$u(x, y) = \frac{1}{2}\left[f(x, y) + g(x, y)\right]$$

substitute $x = (z + \bar{z})/2, y = (z - \bar{z})/2\imath$ to obtain the identity

$$u\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\imath}\right) = \frac{1}{2}\left[f\left(\frac{z+\bar{z}}{2}, \frac{z+\bar{z}}{2\imath}\right) + g\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\imath}\right)\right]. \tag{4.46}$$

Since g is anti-holomorphic, we have $\dfrac{\partial g}{\partial z} = 0$. **Therefore, g is a function of $\bar{z}$ alone.**[6] Hence, upon putting $\bar{z} = 0$ in the above identity, we obtain,

$$u\left(\frac{z}{2}, \frac{z}{2\imath}\right) = \frac{1}{2}\left[f(z) + g(0)\right]. \tag{4.47}$$

However, $g(0) = f(0)$. Since, we want $\Re(f(0)) = u(0,0)$, we can choose $f(0) = u(0,0)$ itself, and obtain

$$\boxed{\textbf{A Magic Formula:}\quad f(z) = 2u\left(\frac{z}{2}, \frac{z}{2\imath}\right) - u(0,0).} \tag{4.48}$$

Of course, we can add a constant imaginary number to f, if we like.

Remark 4.9.2 There is one other small point in the above considerations which we should not ignore, viz., the domain of definition of u was assumed to contain the origin. This can always be arranged by making a suitable translational change of co-ordinates.

Example 4.9.2 Let us find the complex differentiable function f such that $\Re(f) = u$ in each of the following cases using the magic formula (4.48).

(i) $u = x^2 - y^2$. Then $f(z) = 2u(z/2, z/2\imath) - u(0,0) = z^2$.

(ii) $u = xy$. So $f(z) = 2u\left(\dfrac{z}{2}, \dfrac{z}{2\imath}\right) - u(0,0) = \dfrac{z^2}{2\imath}$.

(iii) $u = \dfrac{y}{x^2+y^2}$.

Observe that u is not defined at 0. So we shift the origin at a convenient point say $z = 1$, and consider the function $u_1(x,y) = \dfrac{y}{(x-1)^2+y^2}$. Then,

$$f_1(z) = 2\frac{z/2\imath}{(z/2-1)^2 + (z/2\imath)^2} = \frac{\imath z}{z-1}.$$

Now to obtain the required function we have to shift the origin back to 0. So we have, $f(z) = f_1(z+1) = \imath(z+1)/z$.

[6] This is the only statement which is not proved or explained properly here. Interested reader may refer to section 3 of Chapter 4 in [Cartan].

As an easy consequence of theorem 4.9.1, we shall now prove that harmonicity is preserved under conformal transformations. Indeed, we do not need the full force of conformality here.

Theorem 4.9.2 *Suppose $f : \Omega_1 \to \Omega_2$ is either a holomorphic function or an anti-holomorphic function and $\phi : \Omega_2 \to \mathbb{R}$ be a harmonic function. Then $\psi = \phi \circ f$ is harmonic.*

Proof: Take $z_1 \in \Omega_1$ and let $z_2 = f(z_1)$. By continuity of f, there exist open discs $B_j \subset \Omega_j, j = 1, 2$ around z_j respectively such that $f(B_1) \subset B_2$. Choose a conjugate $\hat{\phi}$ to ϕ in B_2 so that $g = \phi + \imath\hat{\phi}$ is holomorphic. Then $g \circ f$ is either holomorphic or anti-holomorphic. In either case, its real part is harmonic. But $\Re(g \circ f) = \Re(g) \circ f = \psi$. ♠

Theorem 4.9.3 Mean Value Property : *Let u be harmonic in a domain Ω and $B_r(z_0) \subset \Omega$. Then*

$$u(z_0) = \frac{1}{2\pi} \int_{|z-z_0|=r} u(z)\, d(\arg(z - z_0)) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{\imath\theta})\, d\theta. \qquad (4.49)$$

Proof: Let v be a harmonic conjugate of u on $B_r(z_0)$ so that $f = u + \imath v$ is holomorphic. By Cauchy's integral formula we have

$$f(z_0) = \frac{1}{2\pi\imath} \int_{|z-z_0|=r} \frac{f(z)}{z - z_0} dz = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + re^{\imath\theta}) d\theta.$$

Upon equating real parts on either side, we get (4.49). ♠

Theorem 4.9.4 Maximum Principle *Let u be a non constant harmonic function in a domain Ω. Then u does not attain its maximum or minimum in Ω.*

Proof: The proof is similar to that of theorem 4.8.1 and simpler. Nevertheless let us write it down, this time for the minimum, for a change.

Let A be the set of all points $a \in \Omega$ such that $u(a) \leq u(z)$ for all $z \in \Omega$. We want to show that $A = \emptyset$. Assume that $A \neq \emptyset$. Let $z_0 \in A$ and $u(z_0) = k$. Then clearly

$$A = \{z \in \Omega \ :\ u(z) = k\}$$

Therefore A is a closed subset of Ω. We shall show that A is also open in Ω. Then from theorem 1.6.2, it follows that $A = \Omega$. This would mean that u is a constant which contradicts the hypothesis.

Pick any $w \in A$ and choose a small disc $B_r(w)$ around w contained in Ω. Then by the mean value property, for $0 < s < r$, we have,

$$k = u(w) = \frac{1}{2\pi}\int_0^{2\pi} u(w + se^{\imath\theta})d\theta \geq \frac{1}{2\pi}\int_0^{2\pi} k\, d\theta = k.$$

Therefore, it follows that

$$\int_0^{2\pi}[u(w + se^{\imath\theta}) - k]d\theta = 0.$$

The integrand is a continuous non negative function. Therefore, it vanishes identically. Thus $u(w + se^{\imath\theta}) = k$ for all $0 \leq \theta \leq 2\pi$. Since this is true for $0 \leq s < r$, we get $B_r(w) \subset A$. Thus, we have proved that A is open. This completes the proof that u does not attain its minimum inside Ω.

By considering the above discussion for the harmonic function $-u$, it follows that u does not attain its maximum either inside Ω. ♠

Remark 4.9.3 Suppose now that $\bar{\Omega}$ is a closed and bounded subset of $\mathbb{C}$ and u is a continuous function on $\bar{\Omega}$ and harmonic in the interior of Ω. Then its extreme values are definitely attained in $\bar{\Omega}$ and by the above theorem these are not in Ω. Therefore these extreme values must occur on the boundary. As an immediate consequence of this we have:

Theorem 4.9.5 *Let f, g be any two continuous functions on a closed and bounded subset D of $\mathbb{C}$ and harmonic in the interior of D. Suppose $f = g$ on the boundary of D. Then $f = g$.*

Proof: Apply maximum principle to the harmonic function $f - g$. Since it vanishes on the boundary, its maximum and minimum are both equal to zero. Therefore $f - g$ is identically zero. ♠

Remark 4.9.4 The above result says that a harmonic function u is completely determined by its value on the boundary of a closed and bounded region. This is going to be extremely useful, if we can actually determine u at all points of the interior. For any general domain, finding a harmonic function which extends a given function on the boundary or part of the boundary of the domain to the whole of the domain goes under the name **Dirichlet's Problem**. Its solution in specific cases has already lead to very important developments. We shall discuss this in greater detail in a latter chapter. Here, let us consider a few simple cases of practical importance. The very first one is the most simple viz., where D is a closed disc. The following result may be viewed as a generalization of Mean Value Theorem.

Poisson Integral Formula: For a point z_0 such that $|z_0| < R, R > 0$, consider the flts

$$T(w) = \frac{R^2(w+z_0)}{R^2+\bar{z}_0 w}; S(z) = \frac{R^2(z-z_0)}{R^2-\bar{z}_0 z}.$$

Check that T, S are inverses of each other, T maps the disc $|z| < R$ onto itself and maps 0 to z_0. Now the function $u \circ T$ is again harmonic on the disc $|z| \le R$ and by (4.49), we have

$$u(z_0) = u \circ T(0) = \frac{1}{2\pi}\int_{|w|=R} (u \circ T)(w) d(\arg w)$$

where $w = S(z)$ varies over the unit circle on which integration is being taken in the w-plane. Therefore, $\arg w = -\imath \ln w$ and hence

$$\begin{aligned} d(\arg w) &= -\imath \frac{dw}{w} = -\imath \frac{d(S(z))}{S(z)} dz \\ &= -\imath \frac{R^2(R^2-\bar{z}_0 z) + R^2(z-z_0)\bar{z}_0}{(R^2-\bar{z}_0 z)^2}\left(\frac{R^2-\bar{z}_0 z}{R^2(z-z_0)}\right) dz \\ &= -\imath \frac{R^2-r^2}{(R^2-\bar{z}_0 z)(z-z_0)} \imath z d\theta = \frac{R^2-r^2}{|z-z_0|^2} d\theta \end{aligned}$$

the last equality being obtained by using the fact $z\bar{z} = R^2$ on the circle of integration $|z| = R$.

Observe that $\Re(z\bar{z}_0) = Rr\cos(\psi)$, where ψ is the angle between the two vectors z, z_0 of modulus R, r respectively. Writing $u(r,\theta)$ for $u(re^{\imath\theta})$, we have proved:

Theorem 4.9.6 Poisson Integral Formula: *Let $u(r,\phi)$ be a harmonic function on an open set containing the closed disc $|z| \le R$. Then*

$$\boxed{u(r,\phi) = \frac{1}{2\pi}\int_0^{2\pi} u(R,\theta)\frac{R^2-r^2}{R^2-2Rr\cos(\theta-\phi)+r^2}\, d\theta} \tag{4.50}$$

for all $0 \le r < R, 0 \le \phi \le 2\pi$.

Remark 4.9.5

(i) Putting $z = re^{\imath\psi}$ formula (4.50) is equivalent to

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{R^2-|z|^2}{|Re^{\imath\theta}-z|^2} u(Re^{\imath\theta}) d\theta, \quad |z| < R. \tag{4.51}$$

(ii) **Schwarz's formula for harmonic conjugate**: Put

$$f(z) = \frac{1}{2\pi\imath}\int_{|w|=R} \frac{w+z}{w-z} u(w) \frac{dw}{w}. \tag{4.52}$$

Then we know that f is a holomorphic function. Use Exercise 1.3.4 and (4.50) to see that $\Re(f(z)) = u(z)$. This immediately gives us a formula for the conjugate v of u, viz.,

$$v(z) = \Im\left(\frac{1}{2\pi\imath}\int_{|w|=R}\frac{w+z}{w-z}u(w)\frac{dw}{w}\right) + a, \tag{4.53}$$

a being a constant. Now differentiate (4.52) with respect to z under the integral sign to obtain an expression for the derivatives of an holomorphic function f purely in terms of $\Re(f)$.

$$f'(z) = \frac{1}{2\pi\imath}\int_{|w|=R}\frac{2}{(w-z)^2}u(w)dw. \tag{4.54}$$

(iii) With the mere assumption on u that the integral (4.52) exists, it follows that the differentiation under integral sign holds. Therefore for any integrable function u on the boundary of the disc, formula (4.50) gives us a harmonic function in the interior of the disc. How is this related with the given function? The following theorem of Schwarz completely answers this aspect of Dirichlet's problem:

Theorem 4.9.7 (**Schwarz**) *Let u be a bounded and piece-wise continuous real valued function on the unit circle. Then*

$$P_u(z) := \frac{1}{2\pi\imath}\int_{|w|=1}\Re\left(\frac{w+z}{w-z}\right)\frac{u(z)}{z}dz \tag{4.55}$$

is harmonic in the unit disc $\mathbb{D}$. If u is continuous at $z = z_0$ then

$$\lim_{z\to z_0} P_u(z) = u(z_0). \tag{4.56}$$

Proof: Since $P_u = \Re(f)$, where f is given by (4.52), from the remark 4.9.5(ii), it follows that P_u is harmonic. To see the required property at a boundary point as above we may assume, for the sake of notational simplicity, that $u(z_0) = 0$ and prove that the limit in (4.56) is zero.

Given $\epsilon > 0$, first find an open arc on the boundary of $\mathbb{D}$ containing z_0 such that $|u(z)| < \epsilon/2$ on this arc. Let us denote the closure of this arc by L_1. Let L_2 denote the complementary arc on the circle. Let $u_1 \equiv u$ on L_1 and $\equiv 0$ on L_2. Then $|u_1(z)| \leq \epsilon/2$ for all $z \in \mathbb{S}^1$ and hence by maximum-minimum principle, it follows that

$$|P_{u_1}(z)| < \epsilon/2, \quad \forall z \in \mathbb{D}. \tag{4.57}$$

Let now $u_2 \equiv 0$ on L_1 and $\equiv u$ on L_2. Then $u = u_1 + u_2$, on $\mathbb{S}^1$ and hence, $P_u = P_{u_1} + P_{u_2}$. So, it suffices to show that $|P_{u_2}(z)| < \epsilon/2$, for $|z - z_0| < \delta$ for a suitable $\delta > 0$. Let M be a bound for u on the circle. Since $u_2 = 0$ on L_1, we have,

$$P_{u_2}(z) = \frac{1}{2\pi\imath}\int_{L_2} \frac{w+z}{w-z}\frac{u(z)}{z}dz. \tag{4.58}$$

Now let us choose $r > 0$ such that the closed disc B of radius r around z_0 does not intersect L_2. Consider the continuous function $\alpha(w,z) = \Re\left(\frac{w+z}{w-z}\right)$ on $L_2 \times B$. Observe that $\alpha(w, z_0) = 0$ for all $w \in L_2$. (*arg* α is the angle subtended at w by the segment $[-z, z]$ and when $z = z_0$ this is the diameter.) By uniform continuity, we can find $\delta > 0$ such that for $|z - z_0| < \delta$, implies $|\alpha(w,z)| < \epsilon/2M$.

Since, the length of the arc L_2 is less than 2π, by M-L inequality, (4.58) implies that follows that

$$|P_{u_2}(z)| < \epsilon/2, \quad \forall\ |z - z_0| < \delta,\ z \in \mathbb{D}. \tag{4.59}$$

Combining (4.57) and (4.59) we get

$$|P_u(z) \le |P_{u_1}(z)| + |P_{u_2}(z)| < \epsilon, \quad |z - z_0| < \delta$$

which proves (4.56) and thereby completes the proof. ♠

Combining a number of results that we have seen so far, let us now obtain another important result due to Schwarz. In part, we have seen it in exercises 6-8 of section 4.5. Here we see this in a more direct way and in a stronger form.

Theorem 4.9.8 Schwarz's Reflection Principle: *Let Ω be a domain in $\mathbb{C}$ which is symmetric with respect to conjugation. Let $\Omega^{\pm}$ denote the part of Ω lying in the upper(resp. lower) half of the plane and let $A = \Omega \cap \mathbb{R}$. Let v be a continuous function in $\Omega^+ \cup A$, harmonic in Ω^+ and identically zero on A. Consider the map defined by*

$$V(z) = \begin{cases} v(z), & \text{if } z \in \Omega^+ \cup A, \\ -v(\bar{z}), & \text{if } z \in \Omega^-. \end{cases}$$

Then V is harmonic on Ω.

Proof: Clearly, the two parts defining V coincide on the common domain viz., on A and hence V is continuous. Also, since $z \mapsto \bar{z}$ is angle preserving, it follows that $-v(\bar{z})$ is harmonic in Ω^-. Therefore, it remains to consider points in A. Put

$$P_V(z) = \frac{1}{2\pi}\int_0^{2\pi} \Re\left(\frac{z_0 + r\exp \imath\theta + z}{z_0 + r\exp \imath\theta - z}\right) V(z_0 + re^{\imath\theta})\, d\theta. \tag{4.60}$$

Then by theorem 4.9.7 P_V is harmonic in $B_r(z_0)$ and equals V on the boundary. We claim that $P_V(z) = 0$ for all $z \in A \cap B_r(z_0)$. This will then prove that $P_V = V$ on $B_r(z_0)^+$ since the two harmonic functions have the same limit function on the boundary. The holds on $B_r(z_0)^-$ as well. But then we have proved that $P_V = V$ all over $B_r(z_0)$. Therefore, the function V is harmonic at z_0.

Thus it remains to prove that $P_V(z) = 0$ for $z \in A \cap B_r(z_0)$. Let then $z \in A$. Using the (4.51) we see that

$$P_V(z) = u(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{R^2 - |z|^2}{|Re^{\imath\theta} - z|^2} u(Re^{\imath\theta})d\theta.$$

Now we know that $u(w) = -u(\bar{z})$ and $|w - z|^2 = |\bar{w} - z|^2$ for $z \in A$. Therefore the integrand

$$f(R,\theta) = \frac{R^2 - |z|^2}{|Re^{\imath\theta} - z|^2} u(Re^{\imath\theta})$$

has the property that $f(R, -\theta) = -f(R,\theta)$. Therefore for $z \in A$,

$$\begin{aligned} P_V(z) &= \frac{1}{2\pi}\left(\int_0^{\pi} f(R,\theta)d\theta + \int_{\pi}^{2\pi} f(R,\theta)d\theta\right) \\ &= \frac{1}{2\pi}\left(\int_0^{\pi} f(R,\theta)d\theta + \int_{-\pi}^{-2\pi} f(R,\theta)d\theta\right) \\ &= \frac{1}{2\pi}\left(\int_0^{\pi} f(R,\theta d\theta) - \int_0^{\pi} f(R,\theta d\theta)\right) = 0. \end{aligned}$$

This completes the proof. ♠

Exercise 4.9

1. Show that $u(x,y) = 2x(1-y) + 1$ is harmonic and find the holomorphic function f with $\Re(f) = u$ and its conjugate by integration method as well as by formal method.

2. Show that sum of two harmonic functions is harmonic and scalar multiple of a harmonic function is harmonic.

3. Find the holomorphic function f so that its real part is given by (i) $x^3 - 3xy^2$; (ii) $\sinh x \sin y$; (iii) $\dfrac{\sin y + \cos y}{\cosh x + \sinh x}$; (iv) $\dfrac{\sin x \cos x}{\cos^2 x + \sinh^2 y}$.

4. Show that a real polynomial $p(X,Y)$ is harmonic iff all of its homogeneous parts are harmonic.

5. Discuss the harmonicity of a real homogeneous polynomial $p(X,Y)$ of degree ≤ 2.

6. Let $p(X,Y) = aX^3 + bX^2Y + cXY^2 + dY^3$ be a homogeneous polynomial of degree 3 in X, Y over the reals. i.e., $a, b, c, d \in \mathbb{R}$. Find the most general condition under which p is harmonic and find a conjugate.

7. Let $n \geq 4$. Determine the necessary and sufficient conditions on the coefficients of $p(X,Y) = \sum_{j+k=n, j\neq k} a_{jk}X^jY^k$, for p to be harmonic. Use this to give a direct proof of the fact that if $p(X,Y)$ is harmonic then $2p(z/2, z/2\imath)$ has real part equal to $p(X,Y)$. [This explains to some extent the mysterious looking arguments involved in the derivation of the magic formula (4.48.]

8. Theorem 4.9.1 holds for more general class of domains also, viz., on simply connected domains(see ch. 7). However, this is not true in the most general case. Give an example of a harmonic function on $\mathbb{C} \setminus \{0\}$ which is not the real part of a holomorphic function.

9. If $f(z)$ is holomorphic in a domain D, show that $|f(z)|^2$ is not harmonic unless $f(z)$ is a constant.

10. Suppose $f : \Omega \to \mathbb{C}$ is a harmonic function and $z \mapsto zf(z)$ is also harmonic. Show that f is holomorphic. Can you generalize this?

11. Prove (4.50) in the more general case, viz., when u is harmonic in the interior of the disc and continuous on the boundary.

12. If you try to employ the arguments used in exercise 4.5.6, in proving theorem 4.9.8, what goes wrong? Assume further that v_x, v_y exist and continuous on A also. Then give an alternative proof of theorem 4.9.8 based on Ex. 4.5.6.

4.10 Application to Potential Theory

Laplace equations occur in the study of incompressible fluid flows, steady-state temperatures, electrostatic fields, gravitational fields etc. and the study goes under the general name of Potential Theory. In the 2-dimensional case, we have already seen the close connection of this with holomorphic functions. In this section we shall merely discuss some mathematical aspects of a few elementary examples to illustrate the usefulness of complex analysis, especially that of conformal mappings, in this study.

Example 4.10.1 Potential between parallel plates The problem is to find a harmonic function representing the potential in a region in $\mathbb{R}^3$ lying between two parallel plates the function being equal to some known constants along each plate.

By choosing the coordinates so that the $x-$axis is perpendicular to the two plates, and then restricting the function to the plane $z = c$ the problem is reduced to 2-dimensional case. Thus we now have to find a harmonic function u on an infinite strip say,

$$\{x + \imath y \; : \; a \leq x \leq b\}$$

such that u is a constant for $x = a$ and $x = b$. We can simply make a guess work: Try some functions which are y-independent. This just means that we are looking for the solution of

$$u_{xx} = 0$$

given $u(a)$ and $u(b)$. Therefore, $u(x) = \alpha x + \beta$ where

$$\alpha = \frac{u(b) - u(a)}{b - a}; \quad \beta = \frac{bu(a) - au(b)}{b - a}.$$

As a specific example, we may take $a = 2, b = 5$ and $u(a, 0) = 1, u(b, 0) = 2$. Then $u(x, y, z) = (x + 1)/3$.

Remark 4.10.1 It is not necessary that the function that we have found above is unique in general, for the simple reason that the domain under consideration is not bounded completely by the prescribed boundary. However, if we demand that the potential itself should be bounded, then it can be shown that the above solution is unique.

Example 4.10.2 Find the **potential between two coaxial cylinders** with its values on each of them being some constant. This problem is similar to the above one except that now, we can use polar coordinates and try for u which is independent of θ. The Laplace equation becomes

$$r^2 u_{rr} + r u_r = 0$$

which after simplification and separating the variables, becomes

$$\frac{u_{rr}}{u_r} = \frac{1}{r}; \; u_r = \frac{c}{r}; \; u(r) = a \ln r + b$$

where a, b are determined by the values $u(r_1), u(r_2)$.

Observe that the function u extends beyond the two cylinders. It has a singularity on the axis of the cylinder ($r = 0$). In the two dimensional version, the point $r = 0$ is a singularity for the potential function. Depending whether a is positive or negative, this singularity is referred to as a **source** or a **sink** of the potential function. The absolute value of a is called the strength of the potential. In general, other types of singularities can occur.

Example 4.10.3 Two-dimensional fluid flows Let Ω be a domain in $\mathbb{R}^2$ and let the function $f : \Omega \to \mathbb{R}^2$,

$$f(x, y) = (u(x, y), v(x, y))$$

represent the velocity vector of a 2-dimensional steady state (independent of time) fluid flow. We assume that f has continuous partial derivatives of second order throughout Ω. A smooth curve $\gamma : [a, b] \to \Omega$ is called a **stream line** of the flow if $\gamma'(t)$ is parallel to $f(\gamma(t))$ for all $t \in [a, b]$. The flow is said to be **irrotational** if $u_y = v_x$ everywhere on Ω. It is called **incompressible** if $u_x = -v_y$. (These conditions can be derived rigorously by physical considerations, though here we have opted them as definitions.)

We may view f as a complex valued function of a complex variable. Then, we see that a steady state 2-dimensional fluid flow represented by f is irrotational and incompressible iff $\bar{f}$ satisfies Cauchy Riemann equations. Under the smoothness condition that we have assumed this is equivalent to say that $\bar{f}$ is holomorphic i.e., f is anti-holomorphic.

Now for simplicity, assume that Ω is a convex region (simply connectedness is enough). We have seen that every holomorphic function in a convex region has a primitive. Choose $g : \Omega \to \mathbb{C}$ such that $g' = \bar{f}$. Then g is called the **complex potential** of the flow. It is determined up to an additive constant and can be computed by taking integral of $\bar{f}$ along any contour starting at a fixed point. Write $g = \phi + \imath\psi$. It follows that $\psi_x = v$ and $\psi_y = u$. Verify that any curve along which ψ is a constant, is a stream line. That is why the imaginary part of the complex potential is called the **stream function** of the flow; the real part is called the **potential function.** The curve $\Re(g) = constant$ are called **equipotential lines**. Potential function and stream function are both harmonic. Working with the complex potential rather than the real potential has the advantage of availability of complex function theory.

Let us now consider a few physically interesting examples.

If the complex potential is given by $g_1(z) = z^2$ then the stream lines and equipotential lines are respectively the hyperbolas $xy = c$; and $x^2 - y^2 = c$. Physically this represents the flow around a corner (Fig.23.A).

The potential $g_2(z) = \ln(1 + z)$ represents a flow whose stream lines are radial rays through 1 and equipotential curves are circles with center at 1. This has a ‘source’ at

$z = 1$ from where the flow is emerging. For the potential $-g_2(z)$ the point 1 becomes a 'sink' (Fig.23.B). For the potential $\imath g_2$ the equipotential lines and stream lines are interchanged(Fig.23.C).

The complex potential given by $g_3(z) = \cosh^{-1}(z)$ represents the flow for which the stream lines are hyperbolas with foci at ± 1. This can be interpreted as the flow through a small aperture (Fig.23.D). The flow around a thin plate has elliptical stream lines and is best represented by $g_4(z) = \cos^{-1} z$ (Fig.23.E).

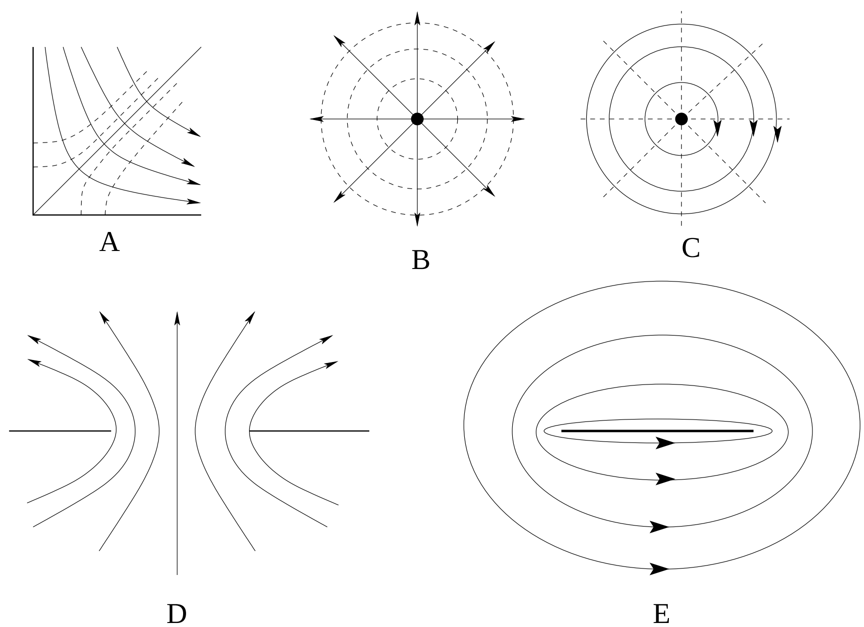


Fig. 23

Example 4.10.4 Potential between two non co-axial Cylinders Assume now that we have to determine a harmonic function on the domain bounded by two circles, one interior to the other, given the value of u on the two boundary components to be constants. Since this problem has been solved for the case when the circles are concentric, we shall try to find a conformal mapping of the given domain with the region between two concentric circles. Without loss of generality, by scaling and translating if necessary we may assume that the outer circle is the unit circle.

We recall that any fractional linear transformation which maps the unit disc onto itself is of the form

$$T(z) = c\frac{z-a}{1-\bar{a}z}$$

with $|c| = 1$ and $|a| < 1$. (See theorem 3.7.5.) Thus, we have two freedoms in the choice of c and a so that T will map the inner circle onto some circle with center 0.

By performing a rotation, we can bring the center of the inner circle to be inside the interval $(0,1)$. This already uses up the freedom in the choice of c. Therefore, we may now assume $c = 1$ and try to fix the value of a in such a way that the inner circle is mapped onto some circle with center 0.

By conformality the diameter of the inner circle should go on to the diameter of $T(C)$. Therefore, x-axis should be mapped onto itself. In particular, a is real. Moreover, the two points α, β of intersection of the inner circle with the x-axis, should be mapped onto $\pm s$ for some $0 < s < 1$ respectively. Therefore, we set up the equation

$$T(\alpha) = -T(\beta)$$

Upon simplification this becomes

$$a^2(\alpha+\beta) - 2a(1+\alpha\beta) + (\alpha+\beta) = 0$$

which is a quadratic equation for a. This admits precisely one real solution a with $|a| < 1$ since $(1+\alpha\beta)^2 > (\alpha+\beta)^2$ for $0 < \alpha < 1$ and $-1 < \beta < 1$.

Example 4.10.5 Steady state temperature T(x, y) in a thin semi-infinite plate $y \geq 0$, whose faces have been insulated and whose edge $y = 0$ is kept at temperature 0 except in the segment $-1 < x < 1$, where the temperature is 1. This translates into the following boundary value problem: Find a harmonic function $T(x,y)$ in the upper half plane such that

$$T(x,0) = \begin{cases} 0, & |x| > 1 \\ 1, & |x| < 1. \end{cases}$$

There are several (slightly) different approaches to this problem, none of which is too obvious.

(A) One function that we are familiar with which assumes two constant values on the two portions of the real axis is $\arg z$. In order to handle the three segments, we should try to find a conformal mapping which will bring the two segments $|x| > 1$ together say onto the positive real axis and map the segment $|x| < 1$ onto the negative real axis. Such a conformal map must map 1 to 0 and -1 to ∞. Lo! we know one such fractional linear transformation:

$$\phi(z) = \frac{z-1}{z+1}.$$

Therefore, $\arg(\phi(z))/\pi$ is the solution that we are looking for.

(B) Geometrically, the points $|x| < 1$ on the real axis are characterized by the property that the angle between the two segments $[-1, x]$ and $[1, x]$ is equal to π where as the points $|x| > 1$ are characterized by the property that the angle between these segments is 0. For points z in the upper-half plane the angle between the segments $[-1, z]$ and $[1, z]$ is the imaginary part of a holomorphic function viz., $\text{Ln}\left(\frac{z-1}{z+1}\right)$ and hence is a harmonic function, which leads to the same solution after dividing by π and simplifying:

$$T(x, y) = \frac{1}{\pi}\tan^{-1}\left(\frac{2y}{x^2 + y^2 - 1}\right)$$

is the answer.

Exercise 4.10

1. Find the flow with a source and a sink of equal strength at $z = a$ and $z = b$ respectively. Show that the equipotential lines are Appolonius's circles. Show that the lines of force are the circles passing through a and b. (See (1.39) and Exercise 1.9.7.)

2. Discuss the above exercise except that now both a, b are sources.

3. Find the stream lines of the following complex potentials and graph them: (i) z^2; (ii) $1/z$; (iii) $z + 1/z$.

4. Graph the configuration of the flow with the complex potential given by $g(z) = -\imath \ln z$. The point $z = 0$ is a singularity which is of the type called a **vortex**.

5. Find the temperature in a thin semi-infinite strip

$$\{y \geq 0, -\pi \leq x \leq \pi\}$$

whose faces have been insulated and the temperature in the two vertical edges is kept at a constant value 0 and in the horizontal edge at a constant value 1.

6. Find the temperature in a thin plate

$$\{(x, y) \ : \ x \geq 0, y \geq 0\}$$

whose faces as well as the portion of the horizontal edge $0 < x < 1$ have been insulated and the temperature is kept at a constant value T_1 on the vertical edge and at T_2 on the rest of the portion of the horizontal edge $x > 1$.

4.11 Miscellaneous Exercises to Ch. 4

1. Given an entire function f having no zeros on a convex domain, show that $f(z) = \exp(g(z))$ for some entire function g.

2. Show that the following integration by parts is valid for complex differentiable functions:

$$\int_\omega f(z)g'(z)\,dz = f(\omega(b))g(\omega(b)) - f(\omega(a))g(\omega(a)) - \int_\omega g(z)f'(z)\,dz,$$

where $\omega : [a, b] \longrightarrow \mathbb{C}$ is any contour.

3. Show that the following integrals taken along the unit circle $|z| = 1$ are all zero.
(a) $\displaystyle\int \frac{e^z}{z-5}\,dz$ (b) $\displaystyle\int \sqrt{z+5}\,dz.$ (c) $\displaystyle\int \frac{\sin z}{z}\,dz$ (d) $\displaystyle\int \frac{\cos z - 1}{z}\,dz.$

4. Evaluate the following integrals along the line segments joining the specified points:
(a) $\displaystyle\int_{-1}^{\imath} \cosh z\,dz.$ (b) $\displaystyle\int_{-\pi\imath}^{\pi\imath} e^z\,dz.$ (c) $\displaystyle\int_{-r}^{r} (z^5 + z)\,dz.$

5. Consider the domain $A = \mathbb{C} \setminus \{z \ : \ z = x < 0\}$. Let ω be any contour in it joining 1 and w. Compute $\displaystyle\int_\omega \ln z\,dz$, where, $\ln z$ is the principle branch of logarithm.

6. Show that $\displaystyle\int_0^\pi e^{a\cos\theta}\cos(a\sin\theta)d\theta = \pi.$

7. Compute $\displaystyle\int_{|z|=2} (z^2+1)^{-1}\,dz.$

8. Find the value of $\displaystyle\int_{|z|=\rho} \frac{|dz|}{|z-a|^2}$, assuming that $|a| \neq \rho$.

9. Find $\displaystyle\int_\omega z^n(z-1)^{-1}\,dz$, where $\omega(t) = 1 + e^{2\pi\imath t},\ 0 \le t \le 1$, and $n \ge 1$.

10. Compute $\displaystyle\int_\omega (z + z^{-1})\,dz$, where ω is the unit circle traced in the counter clockwise sense.

11. $\star$ Show that if X, Y are path connected subspaces such that $X \cap Y$ is non empty, then $X \cup Y$ is path connected.

12. $\star$ Show that if B is a path connected subset of X then B is contained in a path component of X.

13. * Let $G \subset \mathbb{C}$ be open. Show that each path connected component of G is open.

14. Let ω be any closed contour in $\mathbb{C}$ and let a be a point not on this contour. Show that for $n \geq 2$,
$$\int_\omega \frac{dz}{(z-a)^n} = 0.$$

15. Let $p(z)$ be a polynomial, k be a positive integer and C be the circle around a point a and of radius r. Compute $\displaystyle\int_C p(z)(z-a)^{-k}\, dz$.

16. Determine $\displaystyle\int_C (3z^2 + 7z + 1)(z-1)^{-1}\, dz$, where C is the ellipse: $x^2 + 2y^2 = 8$.

17.* Let $\triangle ABC$ be an equilateral triangle in $\mathbb{R}^2$. Start at the midpoint M_1 of AB, join it to the opposite vertex C and trace the line segment M_1C up to the midpoint M_2 of CM_1. Extend BM_2 to meet the side AC at N_2. Let M_3 be the midpoint of CN_2. Trace this segment from M_2 to M_3. Repeat this precess infinitely. Observe that the sequence of points M_j converges to the midpoint of M_0 of BC. Show that this process defines a non rectifiable continuous path.

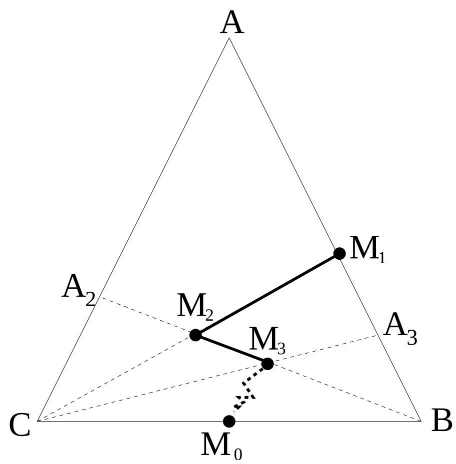


Fig. 24

18. Consider the so called Legendre[7] Polynomials
$$P_n(z) = \frac{1}{n!2^n}\frac{d^n}{dz^n}(z^2-1)^n,\ \ n = 0, 1, 2, \ldots$$

[7]Andrien-Marie Legendre(1752-1833) was a French mathematician who could be compared, to a large extent with Gauss. He contributed to the theory of quadratic reciprocity, did important work in geodesy and theoretical astronomy, studied the attraction of ellipsoids, introduced the so called Legendre functions, formulated the method of least squares and shared several other common interests with Gauss.

Show that for any circle C and for any z in the inside of C, we have,

$$P_n(z) = \frac{1}{2^{n+1}\pi\imath}\int_C \frac{(\xi^2-1)^n}{(\xi-z)^{n+1}}d\xi.$$

Show also that $P_n(1) = 1$ and $P_n(-1) = (-1)^n$.

19. Let f be an anti-holomorphic function. Does f satisfy the maximum modulus principle?

Chapter 5

Zeros and Poles

5.1 Zeros of Holomorphic Functions

Let f be holomorphic in a domain Ω, and let $a \in \Omega$. Assume that all the derivatives of f vanish at a, and $f(a) = 0$. Then from the Taylor series representation, it follows that $f(z) = 0$ for all z in $B_r(a) \subset \Omega$, for some $r > 0$. Putting this in another way, if a holomorphic function is not identically zero in a nbd of a point a, then $f^{(k)}(a) \neq 0$ for some integer $k \geq 0$. In fact, we have,

Theorem 5.1.1 *Let f be a holomorphic function in a domain Ω. Suppose there is a point $a \in \Omega$ such that $f^{(k)}(a) = 0$ for all $k \geq 0$. Then $f \equiv 0$ on Ω.*

Proof: Let $A = \{z \in \Omega : f^{(k)}(z) = 0 \text{ for all } k \geq 0\}$. (Here, $f^{(0)(a) = f(a)}$.) By the hypothesis, A is non-empty. We shall show that A is open as well as closed in Ω. Then by the connectivity of Ω, it follows that $A = \Omega$, and hence, $f \equiv 0$ on Ω.

Let $z_0 \in A$. Choose a disc $B_r(a) \subset \Omega$ on which f is represented by its Taylor's series. It follows that $f(z) = 0$ for all $z \in B_r(a)$. But then it also follows that $f^{(k)}(z) = 0$ for all $z \in B_r(z_0)$ and for all $k \geq 0$. Hence $B_r(0) \subseteq A$. This shows that A is open.

Let now $w \in \Omega$ be a closure point of A. Then there exists a sequence z_n in A such that $z_n \longrightarrow w$. This means that, for each $k \geq 0$, $0 = f^{(k)}(z_n) \longrightarrow f^{(k)}(w)$ and hence $f^{(k)}(w) = 0$. Hence, $w \in A$. This shows that A is closed. This completes the proof of the theorem. ♠

Remark 5.1.1 Observe that the connectivity of the domain plays a mild but logically necessary role here. For, if the function were to be considered on the union of two disjoint open sets, A and B say, we could simply take $f \equiv 0$ on A and any other holomorphic function on B to obtain a counter example. Thus, if we do not assume that Ω is connected, then the conclusion will be that $f \equiv 0$ on the connected component of Ω which contains the point a. In contrast, see the example below for a typical C^∞-function of a real variable with the property $f^{(k)}(0) = 0,\ \forall\, k$, but $f \not\equiv 0$. The above theorem thus leads us to the following definition exclusively for holomorphic functions:

Definition 5.1.1 Let f be a holomorphic function which is not identically zero in a domain Ω. Let $a \in \Omega$. Then the smallest integer $k \geq 0$ such that $f^{(k)}(a) \neq 0$ is called the *order of the zero of f at a.*

Remark 5.1.2 Note that here $f^{(0)}$ just means f. Also *a zero of order zero is not a zero at all!*

Theorem 5.1.2 *Let f be a non zero holomorphic function in a domain Ω and $a \in \Omega$ be a zero of f of order k. Then there is a unique holomorphic function ϕ in a nbd of a such that $\phi(a) \neq 0$ and*

$$f(z) = (z-a)^k \phi(z) \tag{5.1}$$

for all $z \in \Omega$.

Proof: The existence of a function ϕ satisfying (5.1) is a direct consequence of Taylor's expansion (4.35), in which, taking $k = n+1$, putting $f^{(j)}(a) = 0$ for $0 \leq j \leq n+1 = k$, only the last term on the RHS survives. Thus, $f(z) = \phi(z)(z-a)^k$. Now taking the k^{th} derivative we get $\phi(a) \neq 0$. For all points other than $z = a$, (5.1) defines ϕ and hence is unique. ♠

As immediate corollaries, we have the following two theorems:

Theorem 5.1.3 *Let f be a holomorphic function not, identically zero in a domain Ω. Then the zero set of f*

$$\mathcal{Z}_f := \{z \in \Omega \ :\ f(z) = 0\}$$

is an isolated subset of Ω.

Proof: Clearly $\mathcal{Z}_f$ is a closed subset of Ω. Let $a \in \mathcal{Z}_f$. Write f as in (5.1). By continuity of $\phi(z)$, this implies that in a neighborhood of a, ϕ does not vanish and the only zero of $(z-a)^n$ is a. Therefore, in such a nbd of a, a is the only zero of $f(z)$. ♠

Theorem 5.1.4 Identity Theorem: *Let f and g be holomorphic functions on a domain Ω. Suppose $K \subset \Omega$ is such that for every $z \in K$, $f(z) = g(z)$ and K has a limit point in Ω. Then $f \equiv g$ on Ω.*

Proof: For the function $f - g$, the set K happens to be a subset of the set of all zeros. Since this set has a limit point, it follows that the set of all zeros of $f - g$ is not an isolated set. Hence, by the above theorem, $f - g \equiv 0$ on Ω. ♠

Remark 5.1.3 The above results may appear in somewhat different wordings. For instance:
(i) Two holomorphic functions agreeing on an open disc, will have to agree on the whole domain containing the disc.
(ii) Two holomorphic functions which agree on an arc which is not a single point will agree on the whole domain containing the arc.
(iii) Two holomorphic functions which agree on a sequence of distinct points z_n which is convergent to say w, will have to agree on the whole domain containing w.
(iv) A holo(=**the whole**)morphic function on a region is 'completely' determined even if we know it on a very small part. This perhaps justifies the name.
(v) This does not necessarily mean that we can effectively compute its value everywhere, in any of the above situations. In contrast, by Cauchy's integral formula, we could actually know the value of a holomorphic function inside a disc, the moment we know it on the boundary circle.

Example 5.1.1 The structure of the set $\mathcal{H}(\Omega)$ of all holomorphic functions on a domain Ω itself is an interesting topic of study. One can add and multiply any two elements of $\mathcal{H}(\Omega)$ to get another one:

$$(f+g)(z) = f(z) + g(z); \quad (fg)(z) = f(z)g(z), z \in \Omega.$$

The standard properties such as commutativity, associativity, distributivity etc. are verified in a straight forward manner. This makes $\mathcal{H}(\Omega)$ into an algebraic object called commutative ring. Note that the constant functions 0 and 1 respectively play the role of additive and multiplicative identity. The important point now is that if $f, g \in \mathcal{H}(\Omega)$ are such that $fg \equiv 0$ then $f \equiv 0$ or $g \equiv 0$. This is the property that makes a commutative ring into an integral domain. To see this, we simply apply the theorem above. If neither f nor g is identically zero, then the set of zeros of each one of them would be an isolated set. Hence, their union would also be an isolated set. But this is the set of zeros of the product fg which is identically zero. Hence this set is also equal to the whole of Ω which is absurd, since, Ω is an open set. The ring $\mathcal{H}(\Omega)$ was the subject of study much before the advent of modern algebra. Perhaps, even the terminology 'integral domain' has its root here.

In contrast, if we consider C^∞ functions, such is not the case. For simplicity, we can now take our domain to be an interval (this can easily be modified for a rectangle for example) and give a counter example. First consider the following function:

$$f(t) = \begin{cases} e^{-1/t} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0. \end{cases} \tag{5.2}$$

It can be easily seen that f has continuous derivatives of all order. (The only point to be worried about is at the origin. Differentiate the function on the positive interval and take limit as $t \to 0^+$.) All the derivatives at 0 vanish. Yet the function is not identically zero. So, that is a counter example for the theorem 5.1.1, in case of C^∞ functions. Now take $g(t) = f(-t)$. Then we see that g is also smooth and $fg \equiv 0$. Thus $C^\infty(\mathbb{R})$ is a commutative ring but not an integral domain.

Exercise 5.1

1. **L'Hospital's rule** Let f, g be holomorphic in a nbd of a and let a be a zero of order k, l respectively of f, g. Show that $\lim\limits_{z\to a} \dfrac{f(z)}{g(z)}$ exists iff $k \geq l$ and in that case this limit is equal to
$$\lim_{z\to a} \frac{f'(z)}{g'(z)} = \cdots = \lim_{z\to a} \frac{f^{l-1}(z)}{g^{l-1}(z)} = \frac{f^{(l)}(z)}{g^{(l)}(z)}.$$
[Hint: theorem 5.1.2.]

2. * Show that any discrete subset of $\mathbb{C}$ is countable. [Hint: If $\mathcal{U}$ is the countable family of all open balls $B_r(z)$, where $z \in \mathbb{Q} \times \mathbb{Q}$ and $r \in \mathbb{Q}^+$, then we get an injective function $A \to \mathcal{U}$.]

3. * There is no direct analogue of theorem 5.1.2 for harmonic functions. However, there is an analogue of theorem 5.1.3, though somewhat weaker. viz., two harmonic functions which agree on a nonempty open subset agree on the whole domain. Equivalently, prove that if a harmonic function on a domain vanishes on a non empty open set then it vanishes on the entire domain. [Hint: First prove this on a convex domain using theorem 4.9.1 and theorem 5.1.3. For an arbitrary domain, join two arbitrary points by a path and cover this path by interlaced discs.]

5.2 Open Mapping Theorem

The following theorem further emphesises the contrast bewteen the geometric behaviour of a holomorphic function and that of a merely a smooth function of two varaibles.

Theorem 5.2.1 Open Mapping Theorem *A non-constant holomorphic function on an open set is an* open mapping.

Proof: Let $f : \Omega \longrightarrow \mathbb{C}$ be a holomorphic function and let U be any open subset of Ω. We must show that $f(U)$ is an open set in $\mathbb{C}$. Let $w_0 \in f(U)$ be any arbitrary point, say, $w_0 = f(z_0)$, $z_0 \in U$. For simplicity, we may assume that $z_0 = 0 = w_0$ and then we have to show that $f(U)$ contains a neighbourhood of 0.

Let $\epsilon > 0$ be such that $V := B_\epsilon(0) \subset U$ and $f(z) \neq 0$ for $|z| = \epsilon$. Such a choice is possible because the set of zeros of f is discrete. It follows that $0 < \delta := inf\{|f(z)| \ : \ |z| = \epsilon\}$. Suppose $w \in \mathbb{C}$ is such that $|w| < \delta$ and $w \notin f(V)$. Put $g(z) = f(z) - w, z \in V$. Then g is holomorphic on V and does not vanish. Therefore $1/g$ is holomorphic in V and by maximum principle it follows that

$$|w|^{-1} = |g(0)^{-1}| < sup\{\frac{1}{|f(z) - w| \ : \ z \in \partial V} = \frac{1}{\delta - |w|}.$$

It follows that $|w| > \delta/2$. We conclude that $B_{\delta/2}(0) \subset f(V)$. This completes the proof. ♠

Remark 5.2.1 Thus it is clear that if $f : \Omega \longrightarrow \mathbb{C}$ is a non constant holomorphic mapping on a non empty open set, then $f(\Omega)$ cannot be contained in any contour.

As a corollary, we can now drop the continuity condition on f^{-1}, as well as the condition of non vanishing of f' in lemma 2.4.1.

Theorem 5.2.2 Branch Theorem: *Let f be a holomorphic in an open set Ω and be injective. Then $f^{-1} : f(\Omega) \longrightarrow \Omega$ is holomorphic.*

Proof: Given $z_0 \in \Omega$, it follows from theorem 7.2.1, that $f'(z_0) \neq 0$. From theorem 7.2.2, it follows that f is an open mapping. Since f is injective, this is the same as saying $f^{-1} : f(\Omega) \longrightarrow \Omega$ is continuous. The rest of the proof is as in the lemma 2.4.1. ♠

For sharper results, you have wait till chapter 7.

Exercise 5.2 ??

5.3 Singularities

Let Ω be a domain in $\mathbb{C}$. If $f(z)$ is a function on a subset of Ω then the points at which f is not defined or those points at which f is not holomorphic are referred to as singularities of f. We shall restrict ourselves to the study of isolated singularities only.

Definition 5.3.1 A point $z \in \Omega$ is called an *isolated singularity* of f if f is defined and holomorphic in a neighborhood of z except perhaps at z.

Remark 5.3.1 In obtaining Cauchy's theorem and integral formulae, the very first step was Cauchy-Goursat theorem 4.3.1 in which we began with the hypothesis that the function under consideration is holomorphic throughout a domain Ω. Later on we weakened this hypothesis to include those functions f which are continuous in Ω and holomorphic on $\Omega \setminus A$, where A is a discrete subset (theorem 4.3.2). This was very crucial for us in that, it was needed in obtaining Cauchy's integral formulae, since, even if we begin with a holomorphic function f, we had to apply the theorem to the function of the form

$$\frac{f(z) - f(z_0)}{z - z_0}$$

which apparently, is holomorphic in $\Omega \setminus \{z_0\}$ and we do not (yet) know whether it is holomorphic at z_0 or not. Of course, we know how to make it continuous at z_0, viz, by taking the limit which turns out to be $f'(z_0)$.

Choose a small disc $D \subset \Omega$ around a given point $a \in A$ such that $D \cap A = \{a\}$. Proposition 4.1.3 tells us that the function defined by

$$\frac{1}{2\pi\imath} \int_{\partial D} \frac{f(\xi)}{\xi - z} d\xi \tag{5.3}$$

is holomorphic in the interior of D with the mere assumption that $f(\xi)$ is continuous on ∂D. Now Cauchy's theorem says that this function is equal to $f(z)$ at all points of D except possibly at a. Therefore, if we re-define $f(a)$ by the formula

$$\frac{1}{2\pi\imath} \int_{\partial D} \frac{f(\xi)}{\xi - a} d\xi$$

then f is holomorphic on the whole of D.

This way, taking points of A one by one, we can extend f defined and holomorphic on $\Omega \setminus A$ to the whole of Ω, as a holomorphic function. We shall summarize this discussion in the following theorem, in a precise form.

Theorem 5.3.1 Riemann's Removable Singularity *Let Ω be a region, $a \in \Omega$ be any point. Suppose f is holomorphic in $\Omega \setminus \{a\}$. Then the following conditions are equivalent:*
(a) *There is a holomorphic function $\hat{f}$ on Ω such that $\hat{f}(z)f(z), z \in \Omega \setminus \{a\}$, i.e., a is a removable singularity of f.*
(b) $\lim_{z\to a} f(z)$ *exists.*
(c) *f is bounded in a nbd of a.*
(d) $\lim_{z\to a}(z-a)f(z) = 0$.

Proof: (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) are obvious. Also, we know that the hard work, if at all is in proving (d) $\Longrightarrow$ (a). In the above discussion, we have seen a proof of (b) $\Longrightarrow$ (a) since the condition (b) allows us to extend f continuously at a.

We shall now prove (d) $\Longrightarrow$ (b). Consider the function $g(z) = (z-a)f(z)$. Then g satisfies (b) and hence can be extended holomorphically to a function $\hat{g}$ on Ω. In fact $\hat{g}(a) = 0$. Suppose the order of this zero is $k(\geq 1)$. Then there exists a holomorphic function h in Ω such that $\hat{g}(z) = (z-a)^k h(z)$. Therefore for all $z\Omega \setminus \{a\}$, we have $f(z) = (z-a)^{k-1}h(z)$. But the RHS is holomorphic in Ω and hence (b) follows. ♠.

Remark 5.3.2 Thus the only essential ftway a removable singularity a can arise is by taking a genuine holomorphic function f around this point and then brutally redefining its value to be something else only at a or merely pretending as if f is not defined at a. In practice however, it often arises when we divide on holomorphic function with another; some of the points at which both functions vanish may now become removable singularities. A typical example is $\frac{\sin z}{z}$ which is discussed in detail in example 5.3.7.

Example 5.3.1 Non-existence of the n^{th} root function: Let Ω be any disc around 0. We shall show that there is no single valued holomorphic branch of $\sqrt[n]{z}$ defined throughout the punctured disc $\Omega' = \Omega \setminus \{0\}$ for $n \geq 2$. Assuming g to be such a function we observe that g is bounded in Ω' and hence has a removable singularity at 0. Hence, we can extend g to the whole of Ω so that $g^n(z) = z$ on the whole of Ω. Now differentiate this identity to obtain $ng^{n-1}(z)g'(z) = 1$ for all $z \in \Omega$. But, since $g^n(0) = 0$, we have, $g^{n-1}(0) = 0$. Hence, plugging in $z = 0$ in the last equation, we get $0 = 1$ which is absurd.

Example 5.3.2

1. If $p(z)$ is a polynomial, then $1/p(z)$ has all its singularities isolated and these are nothing but the zeros of $p(z)$.

2. Since for any holomorphic function f, the zeros of f are isolated, it follows that all the singularities of $1/f$ are isolated.

3. Natural examples of holomorphic functions which have non isolated singularities are branches of logarithmic function and inverse-trigonometric functions. For instance, $\mathrm{Ln}(z)$ has singularities along the negative real axis.

Definition 5.3.2 Let now $z = a$ be an isolated singularity of f. Suppose $f(z) \to \infty$ as $z \to a$. We then say that $f(z)$ has a *pole* at $z = a$, or $z = a$ is a *pole* of $f(z)$. Since $f(z) \to \infty$ as $z \to a$, $\exists\, \delta > 0$ such that $B_\delta(a) \subset \Omega$ and $f(z) \neq 0$ in $B_\delta(a)$. Consider $g(z) = 1/f(z)$ on $B_\delta(a) \setminus \{a\} =: \Omega'$. Then $g(z)$ is holomorphic on Ω' and is bounded. Hence $z = a$ is a removable singularity of $g(z)$. Observe that $g(z) \to 0$ as $z \to a$ and so we are forced to define $g(a) = 0$ in order to obtain a holomorphic extension of g on $B_\delta(a)$. Suppose a is a zero of g of order k. Then we say that a is a *pole of order* k of f. In this case, from remark 5.1.2, it follows that

$$f(z) = (z-a)^{-k} f_k(z) \tag{5.4}$$

for all z in a neighborhood of a, where $f_k(z)$ is holomorphic and $f_k(a) \neq 0$.

The simplest example of a function with a pole at $z = 0$ of order k is $1/z^k$.

Definition 5.3.3 A function which has all its singularities, if any, as poles, is called a *meromorphic* function in Ω.

Remark 5.3.3 Observe that the poles of a meromorphic function are required to be isolated. Typical examples of meromorphic functions that we came across already are the rational functions. Sums, products and scalar multiples of meromorphic functions are meromorphic. If f and g are non zero meromorphic functions then so is f/g. (Thus, the set of all meromorphic functions on Ω becomes a field, denoted by $\mathcal{M}(\Omega)$.) Further, the zeros of g become poles of f/g, in general. However, if $z = a$ is a common zero of f and g, it becomes a removable singularity of f/g provided the order of the zero of f at a is bigger than or equal to that of g. A typical example of this type is $(\sin z)/z$, discussed in detail in the example below.

Definition 5.3.4 Let f have a pole of order k at $z = a$ and consider $f_k(z) = (z-a)^k f(z)$, and apply the Taylor's expansion:

$$f_k(z) = b'_0 + b'_1(z-a) + \cdots + b'_{k-1}(z-a)^{k-1} + \phi(z)(z-a)^k$$

where ϕ is holomorphic in a nbd of $z = a$. For $z \neq a$, we can divide this expression by $(z-a)^k$ and write $b_1 = b'_{k-1}$, $b_2 = b'_{k-2}, \ldots, b_k = b'_0$, to obtain

$$\boxed{\boxed{f(z) = \frac{b_k}{(z-a)^k} + \frac{b_{k-1}}{(z-a)^{k-1}} + \cdots + \frac{b_1}{(z-a)} + \phi(z).}} \tag{5.5}$$

The sum of terms which involve b_i is called the *principal part* or the *singular part* of $f(z)$ at $z = a$. Observe that f minus its singular part is a holomorphic function.

Further, if we write Taylor's expansion for $\phi(z)$ on the rhs above, what we get is called a *Laurent*[1] *expansion* for $f(z)$ (more about them in section 5.3).

Example 5.3.3 Partial Fractions For a rational function $f = P/Q$, where P and Q have no common factors, the singularities are all poles, and they are precisely the zeros of Q. If a is a zero of Q of order k then in the expression (5.5) for $f(z) = \frac{P(z)}{Q(z)}$, the term $\phi(z)$ on the rhs is again a rational function, being the difference of two rational functions. We also know that it is holomorphic at a and hence has number of poles one less than those for $f(z)$. Repeating this process finitely many times, if $a_1, \ldots, a_k$ are the distinct zeros of Q of respective order $n_1, \ldots, n_k$, we obtain polynomials F_j of degree n_j each having contant term zero, such that

$$\frac{P(z)}{Q(z)} - \sum_{j=1}^{k} F_j\left(\frac{1}{z-a_j}\right)$$

is a rational function which is holomorphic throughout $\mathbb{C}$ and hence is a polynomial function $G(z)$. Thus we obtain an expression

$$\frac{P(z)}{Q(z)} = \sum_{j=1}^{k} F_j\left(\frac{1}{z-a_j}\right) + G(z) \tag{5.6}$$

which is called the *partial fraction* representation of the rational function $f(z)$. Note that if we consider the point at infinity as well into the discussion, the polynomial $G(z)$ is nothing but the singular part of $P(z)/Q(z)$ at ∞ except for the constant term. It can

[1]Pierre Alphonse Laurent (1813-1854) was a French Engineer cum mathematician who proved this theorem around 1843.

be directly computed by carrying out division of $P(z)$ by $Q(z)$ till its degree becomes smaller than that of $Q(z)$. In practice, there are at least two well known methods for computing the polynomials F_j, neither of which can claim to be better than the other except in certain situations. We shall illustrate this, with a number of examples below. In Chapter 8, we shall generalize this to all meromorphic functions.

Remark 5.3.4 The coefficients b_j in (5.5) are given by the integral formula

$$\boxed{b_j = \frac{1}{2\pi\imath}\int_{C_a}(z-a)^{j-1}f(z)\,dz} \tag{5.7}$$

where C_a is a small circle centered at a and oriented counter clockwise. Putting $g(z) = (z-a)^k f(z)$ where k is the order of the pole a, and using Cauchy Integral formula for derivatives, we obtain

$$\boxed{b_j = \tfrac{g^{(k-j)}(a)}{(k-j)!}.} \tag{5.8}$$

Example 5.3.4 For $a \neq b$, consider the partial fraction development

$$\frac{1}{(z-a)^m(z-b)^n} = \sum_{j=1}^{m}\frac{\alpha_j}{(z-a)^j} + \sum_{j=1}^{n}\frac{\beta_j}{(z-b)^j}$$

Put $\phi(z) = \dfrac{1}{(z-a)^m}$; $\psi(z) = \dfrac{1}{(z-b)^n}$. Let C_a be a 'small' circle centered at a. Then from (5.7), we have,

$$\begin{aligned}\alpha_j &= \frac{1}{2\pi\imath}\int_{C_a}\frac{\psi(z)}{(z-a)^{m-j+1}}\,dz\\ &= \frac{\psi^{(m-j)}(a)}{(m-j)!} = (-1)^{m-j}\frac{\binom{m+n-j-1}{m-j}}{(a-b)^{m+n-j}}.\end{aligned}$$

Likewise, we have,

$$\beta_j = (-1)^{n-j}\frac{\binom{m+n-j-1}{n-j}}{(b-a)^{m+n-j}}.$$

Example 5.3.5 Paravartya Sutra Let us obtain the partial fraction development of $f(z) = \frac{z^3}{z^2-1}$ by the school-algebra method. The division of z^3 by z^2-1 yields

$$f(z) = z + \frac{z}{z^2-1}.$$

Since $z^2 - 1 = (z+1)(z-1)$ has ± 1 as simple roots, we seek to find two constats b_1, b_2 such that

$$\frac{z}{z^2-1} = \frac{b_1}{z+1} + \frac{b_2}{z-1}.$$

Upon clearing the denominator, we obtain

$$z = b_1(z-1) + b_2(z+1)$$

Putting $z = 1$ and $z = -1$ yields respectively, $b_1 = b_2 = 1/2$. This is indeed the ancient Indian method called *'Paravartya Sutra'*, which is very effictive when all poles are simple poles. The same result can be arrived at by applying the integral formula (5.7). Verify this.

Example 5.3.6 Here let us take $f(z) = \frac{1}{(z-1)z^2(z+1)^3}$. Of course, $G(z) = 0$ here. Looking at the factors of the denominator, we write

$$f(z) = \frac{a}{z-1} + \frac{b_1}{z} + \frac{b_2}{z^2} + \frac{c_1}{z+1} + \frac{c_2}{(z+1)^2} + \frac{c_3}{(z+1)^3}.$$

By clearing the denominator we get, $1 = az^2(z+1)^3 + b_1(z-1)z(z+1)^3 + b_2(z-1)(z+1)^3$ $+ c_1(z-1)z^2(z+1)^2 + c_2(z-1)z^2(z+1) + c_3(z-1)z^2$.

We can now combine *Paravartya Sutra* above with some other cleverness. Putting $z = 1, 0, -1$ respectively, yields $a = 1/8, b_2 = -1$, and $c_3 = -1/2$. Substituting these values we get polynomial identity of degree 2. By substituting for z any three values other than $-1, 0, 1$ we can now obtain three equations which can be solved for b_1, c_1, c_2. We shall follow another sure-fire method here: Differentiate the above equation once and put $z = 0$ to get

$$0 = \cdots + b_1[(z-1)(z+1)^3] + b_2[(z+1)^3 + 2(z-1)(z+1)^2] + \cdots = -b_1 - 2b_2$$

where we have not bothered to write down the terms which have z as a factor. This gives $b_1 = 2$. Likewise, by ignoring the terms which have $(z+1)$ as a factor we get

$$0 = \cdots + c_2[(z-1)z^2 + \cdots] + c_3[z^2 + 2(-1)z] = -2c_2 + 5c_3$$

and hence $c_2 = -5/4$. Similarly differentiating twice and putting $z = -1$ we get $0 = -4c_1 + 10c_2 - 8c_3$ which gives $c_1 = -21/4$.

Remark 5.3.5

1. Observe that, if $z = a$ is a pole of f then it follows from the above analysis that for some positive integer n, we have, $\lim\limits_{z\to a} |z-a|^n|f(z)| = 0$. In fact, this is so for all n greater than the order of the pole.

2. If f has a pole of order k at $z = a$ and g is holomorphic at $z = a$, then $f + g$ also has a pole of order k at $z = a$.

3. At a removable singularity we could redefine the function so that it becomes holomorphic. At a pole, the situation is only a little worse. If we think of f as taking values in $\widehat{\mathbb{C}}$, then we can assign the value ∞ to f at the pole so that the new function is now continuous. Indeed, in a neighborhood of the pole, $1/f$ becomes complex differentiable. Thus, a meromorphic function f on a domain Ω is indeed a holomorphic function $f : \Omega \longrightarrow \widehat{\mathbb{C}}$. In this sense, not only poles and zeros of a meromorphic functions should be treated on par qualitatively, we can just talk about the solution set of the equation $f(z) = w$ for any $w \in \widehat{\mathbb{C}}$.

4. Following the line of thought in (3) for a meromorphic function, we say an integer k is the *algebraic order of* f *at* a if either a is a zero of order k when $k > 0$ or a is a pole of order $-k$ when $k < 0$. Of course, if $k = 0$ this terminology means that f is holomorphic at a and $f(a) \neq 0$.

   We shall now consider the case, when the singularity is indeed quite *bad.*

Definition 5.3.5 Let $z = a$ be an isolated singularity of $f \not\equiv 0$. It may turn out that $z = a$ is neither a removable singularity nor a pole. Such a singularity is called an *essential singularity.*

Theorem 5.3.2 *Let a be an isolated singularity of f. Then a is an essential singularity iff for* **no** *real number r, the limit $\lim_{z\to a} |(z-a)|^r|f(z)|$ exists.*

Proof: We have already seen that if a is either a removable singularity or a pole, then there exists a positive integer r such that the above limit exists, indeed $r = 1$ in the former case and $r = k + 1$ in the latter case, where k is the order of the pole.

Conversely, assume that we have such real number r. Then for any integer $n > r$, we have $|z-a|^n|f(z)| \longrightarrow 0$. Therefore the function $g(z) = (z-a)^{n-1}f(z)$ has a removable singularity at a. Therefore a is a zero of g of order $m \geq 0$ and we can write $g(z) = (z-a)^m h(z)$, where h is a holomorphic function in a nbd of a and $h(a) \neq 0$. Therefore $f(z) = (z-a)^{m-n+1}h(z)$ for all $z \neq a$ in a nbd of a. This clearly means that $z = a$ is either a removable singularity (when $m \geq n-1$) or a pole (when $m < n-1$). That is, it is not an essential singularity. ♠

The following theorem which can be named as *arbitrary theorem* describes the nature of f near an essential singularity.

Theorem 5.3.3 Casorati[2]-Weierstrass : *Let f be holomorphic in $B_r(a)\setminus\{a\}$ and let a be an essential singularity of f. Then f takes values arbitrarily close to any arbitrary complex number inside any arbitrary neighborhood of a.*

Proof: Given $w \in \mathbb{C}$ and two positive real numbers δ_1, δ_2 we must show that

$$f(B_{\delta_1}(a) \setminus \{a\}) \cap B_{\delta_2}(w) \neq \emptyset.$$

Assuming on the contrary, there exists $w \in \mathbb{C}$ and $\delta_1, \delta_2 > 0$ such that we have $|f(z) - w| \geq \delta_2$ for all $z \in B_{\delta_1}(a) \setminus \{a\}$. Therefore the function $g(z) = \dfrac{1}{f(z) - w}$ is holomorphic and bounded in $B_{\delta_1}(a) \setminus \{a\}$. Therefore a is a removable singularity of g. We can then write $g(z) = (z-a)^k h(z)$, where $k \geq 0$ is some integer, h is holomorphic in $B_{\delta_1}(a)$ and $h(a) \neq 0$. This means $f(z) = w + \dfrac{1}{(z-a)^k h(z)}$ has either a removable singularity or a pole at a. This contradicts the hypothesis. ♠

Remark 5.3.6 This remarkable theorem tells us how wildly a function behaves near an essential singularity. It means that the closure of the image under f of any punctured ball around a is the whole of $\mathbb{C}$. This will not surprise you if you realize that in the study of functions of 1-variable, there are discontinuities of a function at which the function oscillates. A typical example of this is the *topologist's sine-curve* $\{(x, \sin\frac{1}{x}) : x > 0\}$. The curve "approaches' every point on the y-axis between $(0,1)$ and $(0,-1)$. This curve is exploited for various pathological purposes by topologists. The nature of a holomorphic function at an essential singularity is similar, but much wilder. In fact, it is much wilder than Casorati-Weierstrass: *If a is an essential isolated singularity of f, then there exists $w \in \mathbb{C}$ such that for every $\delta > 0$, $f(B_\delta(a)) \supset \mathbb{C} \setminus \{w\}$*. This is called the *'big' Picard[3] Theorem*, which we shall prove in the last chapter.

Example 5.3.7

[2]Falice Casorati(1835-1890) was an Italian Mathematician. He discovered this result in 1868. After eight years, Weierstrass proved this independently.

[3]Charles Emile Picard(1856-1941) a French mathematician chiefly known for his contribution to the theory of analytic functions, existence of solutions of ordinary differential equations, geometry of algebraic surfaces and analytical study of heat, elasticity and electricity.

1. Consider the function $f(z) = \dfrac{\sin z}{z}$, $z \neq 0$. Obviously $z = 0$ is an isolated singularity. We easily see that $\lim\limits_{z \to 0} z f(z) = 0$. Hence, $z = 0$ is a removable singularity. Also we see that $\lim\limits_{z\to 0} f(z) = 1$. So we can define $f(0) = 1$ and make f holomorphic at $z = 0$ also.

2. Let us now consider $g(z) = \dfrac{\text{Ln } (1+z)}{z^2}$, $z \neq 0$. Clearly, g is defined and holomorphic in $B_1(0)\backslash\{0\}$. Observe that $\lim\limits_{z\to 0} g(z) = \infty$, and hence $z = 0$ is a pole of g. To determine the order of the pole, we write Taylor's expansion of Ln $(1+z) = z + z^2\phi(z)$ up to order 2 terms and divide by z^2 to see that $g(z) = \dfrac{1}{z} + \phi(z)$. Thus we conclude that the order of the pole is 1.

3. Let now $h(z) = e^{1/z}$, $z \neq 0$. Then clearly $z = 0$ is an isolated singularity. We observe that, for $r > 0$, and for any positive integer n,
$$r^n h(r) > r^n\left(1 + \frac{1}{r} + \cdots + \frac{1}{n! r^{n+1}}\right) > \frac{1}{n! r}.$$
Therefore, it follows that, for any positive integer n, $\lim\limits_{z\to 0} |z^n h(z)| \neq 0$. From our earlier remark, this implies that $z = 0$ is an essential singularity of h. (This can also be seen by checking that for real y, the quantity $e^{1/\imath y}$ oscillates as $y \longrightarrow 0$.)

Remark 5.3.7

1. The discussion of isolated singularity can be carried out for the point $z = \infty$ as well. To begin with we need that the function is defined and holomorphic in a neighborhood of infinity, i.e., in $|z| > M$ for some sufficiently large M. We say that ∞ is a removable singularity or a pole of f iff
$$\lim_{z\to\infty}\left|\frac{f(z)}{z^n}\right| = 0$$
for some integer n. (If this integer can be chosen to be ≤ 1, then ∞ is a removable singularity, otherwise, it is a pole.) Observe that this is the same as saying that 0 is a removable singularity or a pole for $g(z) = f(1/z)$.

2. With this terminology, as seen in remark 5.3.5.3, we can now consider domains $\Omega \subset \widehat{\mathbb{C}}$ for holomorphic functions with values in $\widehat{\mathbb{C}}$. Note that they are **not** the same as meromorphic functions $\Omega \setminus \{\infty\} \to \mathbb{C}$.

3. The simplest examples are polynomial functions of degree $d \geq 1$. They have a pole only at ∞ and the order of the pole is d. More generally, any rational function

of positive degree d has a pole of order d at ∞; if the degree d is ≤ 0, then it is a removable singularity. Note that if $f, g : \Omega \longrightarrow \widehat{\mathbb{C}}$ are holomorphic functions, then $f+g$ and fg make sense and are both holomorphic functions $\Omega \longrightarrow \widehat{\mathbb{C}}$. In particular, if ∞ is a removable singularity or a pole of f then so it is for pf, where p is a polynomial. Thus, we know that any rational function can be treated as a holomorphic function $\widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$. The following theorem, which can be proved in different ways, is a converse to this.

Theorem 5.3.4 *Let $f : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}$ be a holomorphic function. Then f is a rational function.*

Proof: Recall that this just means that f is a meromorphic function on $\mathbb{C}$ with ∞ as a removable singularity or a pole. So, in a nbd of ∞, f is holomorphic and hence the number of poles of f in $\mathbb{C}$ is bounded. Being a bounded isolated set, the set of all poles of f is finite. [See exercise 5.3.5 and 5.3.6.] Say, $z_1, \ldots, z_m$ are the poles with respective order $k_1, \ldots, k_m$. Put $g(z) = (z-z_1)^{k_1} \cdots (z-z_m)^{k_m} f(z)$. Then g is an entire function which still has a removable singularity or a pole at ∞, and hence $\lim_{z\to\infty} z^{-n} g(z) = 0$ for some integer $n \geq 0$. Being an entire function, g has a power series representation around 0 valid in the whole plane. In particular, we have the Taylor's expansion

$$g(z) = g(0) + g^{(1)}(0)z + \cdots + \frac{g^{(n-1)}(0)}{(n-1)!} z^{n-1} + g_n(z) z^n$$

where g_n is an entire function. It follows that $\lim_{z\to\infty} g_n(z) = 0$. By Liouville's theorem, $g_n \equiv 0$. This means that g is a polynomial and hence f is a rational function. ♠

Corollary 5.3.1 The group of all automorphisms of the Riemann sphere $\widehat{\mathbb{C}}$ consists precisely of all fractional linear transformations.

Proof: We already know that every flt defines an automorphism of $\widehat{\mathbb{C}}$. Conversely, let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be an automorphism. By the above theorem $f = \frac{P}{Q}$, where P, Q are both polynomials without any common factors, say. Let a, b be the leading coefficients of P, Q respectively. Pick up any $w \in \mathbb{C}$ such that $w \neq a/b$. Then if P or Q is of degree bigger than 1, it follows that $P(z) - wQ(z)$ is a polynomial of degree bigger than 1. Therefore it has more than one root say z_1, z_2. But then $f(z_1) = w = f(z_2)$ and hence f is not one-one. Therefore both P and Q are of degree ≤ 1. This means f is a fractional linear transformation. ♠

Corollary 5.3.2 The group of all automorphisms of the plane $\mathbb{C}$ consists precisely affine transfomrations $z \mapsto az + b$, with $a \in \mathbb{C}^*, b \in \mathbb{C}$.

Proof: That the above maps are actually automorphisms of the plane is obvious. We have to prove the converse. So, let $f : \mathbb{C} \to \mathbb{C}$ be a bijective holomorphic mapping such that f^{-1} is also holomorphic. Then first of all by continuity of f^{-1} it follows that if K is a compact subset of $\mathbb{C}$ then $f^{-1}(K)$ is compact. This implies that if $z_n \to \infty$ then $f(z_n) \to infty$. Therefore, ∞ is a pole of f. In particular, f extends to a holomorphic mapping $\hat{f} : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ where $\hat{f}(\infty) = \infty$. Since the same applies to f^{-1} as well it follows that $\hat{f}$ is an automorphism of $\widehat{\mathbb{C}}$. From the corollary above, we conclude that $\hat{f}$ is a fractional linear transformation. Since this takes $\mathbb{C}$ to $\mathbb{C}$ it is actually an affine linear transformation, i.e., $f(z) = az + b$ for some $a, b \in \mathbb{C}$. Clearly $a \neq 0$. ♠

Exercise 5.3

1. Let $f = p/q$ be a rational function. Put $d = \deg p - \deg q$. Prove the following:
   (a) If $d > 0$ then ∞ is pole of f of order d.
   (b) If $d < 0$ then ∞ is a zero of order $-d$.
   (c) If $d = 0$ then ∞ is a removable singularity and the value of f at ∞ is equal to a_0/b_0, where a_0 and b_0 are the leading coefficients of p and q respectively.

2. If f and g are having algebraic order h and k respectively at $z = a$, show that
   (a) $f + g$ has order $\leq \max\{h, k\}$.
   (b) fg has order $h + k$.
   (c) f/g has order $h - k$.
   (d) f' has order $h - 1$, if $h \neq 0$.

3. Let $\mathcal{M}(\Omega)$ denote the set of all meromorphic functions in a domain Ω. Show that $\mathcal{M}(\Omega)$ is a field and contains the integral domain $\mathcal{H}(\Omega)$. [Indeed, $\mathcal{M}(\Omega)$ is the quotient field of $\mathcal{H}(\Omega)$. Trying to prove this leads to the so called '*Mittag-Leffler problem*'. We shall not handle it at this stage.]

4. For each of the following functions determine the nature of the singularity at $z = 0$. Whenever it is a removable singularity, find the value of the function at $z = 0$. Whenever it is a pole find the singular part.

$$(a)\quad \frac{\cos z}{z}; \qquad (b)\quad \frac{\cos z - 1}{z^2}; \qquad (c)\quad \frac{\cos(z^{-1})}{z^{-1}}; \qquad (d)\quad e^{z^{-1}};$$

$$(e)\quad \frac{z^2+1}{z(z-1)}; \qquad (f)\quad (1-e^z)^{-1}; \qquad (g)\quad z \sin\frac{1}{z}; \qquad (h)\quad z^{2n+1}\sin\frac{1}{z}.$$

5. Let $\{z_n\}$ be a sequence of points with w as the limit point. Suppose further that
   (a) $f(z_n) = 0$ for all n or
   (b) f has a pole at z_n for all n.
   Show that w is a singularity of f. Show that in either case, f behaves as described in Casorati-Weierstrass at the point w. [In view of our theme of treating zeros and poles on par, we now detect a defect in our treatment of singularities, i.e, we should have allowed points a of the type (b) also in our treatment, even though it is not an isolated singularity, for, it is an isolated singularity of $1/f$. With such terminology, we would then be able to call the singularity of the above type also as an essential singularity.]

6. Let f be a meromorphic function in a domain Ω. Then show that the set of poles of f is a discrete subset of Ω.

7. Let z_0 be an isolated singularity of f. If $\Re(f)$ or $\Im(f)$ is bounded in some neighborhood of z_0, then show that z_0 is a removable singularity of f.

8. Let ζ denote a primitive n^{th} root of unity. Prove the formula
   $$\frac{1}{z^n - 1} = \sum_{j=1}^{n} \frac{\zeta^r}{z - \zeta^r}. \tag{5.9}$$
   (Hint: See 1.9.4)

9. Let ζ be a primitive n^{th} root of unity:
   Prove the formula:
   $$\sum_{r=1}^{n-1} \frac{1}{1 - \zeta^r} = \frac{n-1}{2}. \tag{5.10}$$

10. Use (5.10) to prove the following partial fraction formula:
    $$\frac{1}{(z^n - 1)^2} = \frac{1}{n^2}\left(\sum_{r=0}^{n-1} \frac{\zeta^{2r}}{(z - \zeta^r)^2}\right) + \frac{1-n}{n^2}\left(\frac{\zeta^r}{z - \zeta^r}\right). \tag{5.11}$$

11. * Obtain partial fraction development of $\dfrac{1}{(z^n - 1)^m}$, where m, n are arbitrary positive integers.

5.4 Laurent Series

A holomorphic function around $z = a$ has a series representation with non negatives powers of $(z-a)$. On the other if $z = a$ is a pole then we have Laurent expansion in which negative powers of $(z-a)$ occur but are finitely many. What happens at an essential singularity? The natural answer seems to be that the number non zero negative poewers is infinite. Indeed, this turns out to be the right answer as we shall see now.

We begin with a few easy lemmas:

Lemma 5.4.1 Let $S(X) = \sum_n a_n t^n$ be a formal power series with radius of convergence, $0 < \rho \leq \infty$. Then $f(z) := S(1/z)$ for $|z| > 1/\rho$, defines a holomorphic function.

Proof: We know that the power series S defines a holomorphic function $g(w) = S(w)$ in $|w| < \rho$. On the other hand the function $z \mapsto \frac{1}{z}$ is a holomorphic function on $\mathbb{C}^*$ and takes the domain $|z| > 1/\rho$ inside $|w| = |1/z| < \rho$. Taking the composite $z \mapsto 1/z \mapsto S(1/z)$, we get the result. ♠.

Thus a series of the form $\sum_{n<0} b_n z^n$ can be thought of as a power series in $1/z$. Recall that any holomorphic function is a disc is represented by a single convergent power series. Notice that the situation here is similar, the function $S(1/z)$ is represented throughout $|z| > 1/\rho$ by a single series. This is the reason we should pay a little more attention to this case. So, let us make a definition.

Definition 5.4.1 *By a Laurent series with center z_0, we mean a sum of the form*

$$\sum_{n<0} b_n(z-z_0)^n + \sum_{n\geq 0} a_n(z-z_0)^n.$$

A Laurent series is said to be convergent at a point z if both the series above are convergent at z. If this happens for a function in a domain Ω then the function which is the sum of these two series is said to have a 'Laurent series representation'.

In what follows, we shall always consider Laurent series with center $z_0 = 0$. Results for the general case would follow if we merely replace z by $z - z_0$.

For $0 \leq r_1 < r_2 \leq \infty$ and $a \in \mathbb{C}$ introduce the notation for the annular domain

$$A(a; r_1, r_2) = \{z \in \mathbb{C} \ : \ r_1 < |z-a| < r_2\}. \tag{5.12}$$

In particular, let $A(r_1, r_2) = A(0; r_1, r_2)$.

Suppose that $S(t) = \sum_{n>0} b_{-n}t^n$ has radius of convergence $\rho = 1/r$, and the series $T(t) = \sum_{n\geq 0} a_n t^n$ has radius of convergence R. Then it follows that $f(z) = S(1/z) + T(z)$ is holomorphic in the annular domain $r < |z| < R$, which is the common region of convergence of both the series. In the following theorem, we will prove the converse to this result.

Lemma 5.4.2 Let Ω be a holomorphic function on $A(0; r_1, r_2)$. Then the function $h : r \mapsto \int_{|z|=r} f(z)dz$ is a constant on the interval (r_1, r_2).

Proof: Consider, $h(r) = \int_0^{2\pi} f(re^{\imath\theta})re^{\imath\theta}\imath\, d\theta = \imath \int_0^{2\pi} g(re^{\imath\theta})\, d\theta$, where $g(z) = f(z)z$. Now for any holomorphic function g, we have

$$\frac{\partial g}{\partial \theta} = g'(re^{\imath\theta})re^{\imath\theta}\imath; \quad \frac{\partial g}{\partial r} = g'(re^{\imath\theta})e^{\imath\theta}.$$

Since g is holomorphic, by differentiating under the integral sign, we have,

$$\frac{dh}{dr} = \imath \int_0^{2\pi} g'(re^{\imath\theta})e^{\imath\theta}\, d\theta = \frac{1}{r}\int_0^{2\pi} \frac{\partial g}{\partial \theta}\, d\theta = \frac{1}{r}[g(r) - g(r)] = 0. \qquad \spadesuit$$

Remark 5.4.1 This lemma can be also derived directly from Cauhcy's Theorem II-version 4.4.2.

Lemma 5.4.3 Let Ω be a holomorphic function on $A = (r_1, r_2)$. Let $r_1 < \rho_1 < \rho_2 < r_2$. Then for $\rho_1 < |z| < \rho_2$, we have

$$\boxed{f(z) = \frac{1}{2\pi\imath}\int_{|w|=\rho_2} \frac{f(w)}{w-z}dw - \frac{1}{2\pi\imath}\int_{|w|=\rho_1} \frac{f(w)}{w-z}dw.} \qquad (5.13)$$

Proof: Define $g : A(r_1, r_2) \longrightarrow \mathbb{C}$ by

$$g(w) = \begin{cases} \dfrac{f(w) - f(z)}{w - z}, & w \neq z \\ f'(z), & w = z. \end{cases}$$

Then g is holomorphic in A and hence, by lemma 5.4.2,

$$\int_{|w|=\rho_2} g(w)dw = \int_{|w|=\rho_1} g(w)dw.$$

We also have

$$\int_{|w|=\rho} \frac{dw}{w-z} = \begin{cases} 2\pi\imath, & |z| < \rho, \\ 0, & |z| > \rho. \end{cases}$$

(See example 4.20). Therefore,

$$\int_{|w|=\rho_2} \frac{f(w)}{w-z}dw - 2\pi\imath f(z) = \int_{|w|=\rho_1} \frac{f(w)}{w-z}dw$$

which proves (5.13).

Theorem 5.4.1 Laurent Series : *Let $0 \le r_1 < r_2 \le \infty$, and let f be holomorphic in the annulus $A(r_1, r_2)$. Then there exists a unique Laurent series, $\sum_{-\infty}^{\infty} c_n z^n$ which is uniformly and absolutely convergent to f on every closed and bounded subset of A. Moreover, the coefficients c_m are given by*

$$c_m := \frac{1}{2\pi\imath}\int_{|z|=r} \frac{f(w)}{w^{m+1}}dw \tag{5.14}$$

for any $r_1 < r < r_2$ and for all $m \in \mathbb{Z}$.

Proof: First we shall prove the uniqueness part. Assuming that $f(z) = \sum_{-\infty}^{\infty} c_n z^n$ is uniformly and absolutely convergent on every compact subset of A, we shall show that 5.14 holds: Let r be such that $r_1 < r < r_2$ and C be the circle of radius r. Then

$$c_m = \frac{c_m}{2\pi\imath}\int_C \frac{dw}{w} = \frac{c_m}{2\pi\imath}\int_C \frac{w^m}{w^{m+1}}dw$$

$$= \frac{1}{2\pi\imath}\int_C \frac{\sum_{-\infty}^{\infty} c_n w^n}{w^{m+1}}dw = \frac{1}{2\pi\imath}\int_C \frac{f(w)}{w^{m+1}}dw.$$

Here the third equality is valid because term-by-term integration is valid and all integrals except the m^{th} one vanish. This shows the uniqueness. It also tells us what to do in order to prove the existence i.e., taking c_m as given above, we have to prove that the corresponding Laurent series converges to $f(z)$ in Ω.

The first thing to observe is that by lemma 5.4.2, c_m defined as above does not depend upon the choice of r.

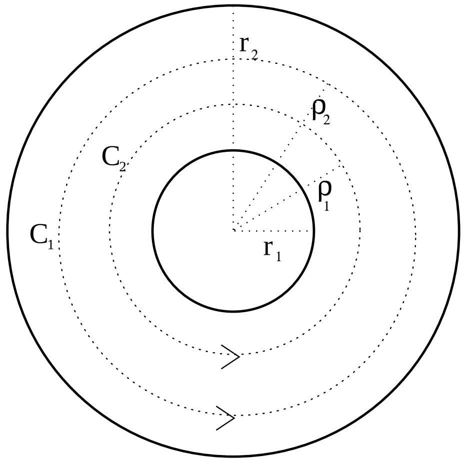


Fig. 25

Now fix ρ_1 and ρ_2 as above, and let $|z| < \rho_2$. Then for $|w| = \rho_2$, we have,

$$\frac{1}{w-z} = \frac{1}{w}\left(\frac{1}{1-z/w}\right) = \frac{1}{w}\left(1 + \frac{z}{w} + \frac{z^2}{w^2} + \cdots +\right) = \sum_0^\infty \frac{z^n}{w^{n+1}}$$

and the series is uniformly convergent on compact subsets of $B_{\rho_2}(0)$. Indeed, the convergence is uniform on compact subsets, with respect to both the variables z and w. Therefore, we can multiply both sides by $f(w)$ and then integrate term-by-term on the circle C_2, to obtain,

$$\frac{1}{2\pi\imath}\int_{C_2}\frac{f(w)}{w-z}\,dw = \sum_0^\infty c_n z^n. \tag{5.15}$$

Similarly, for $|w| = \rho_1$, and $|z| > \rho_1$, we have,

$$\frac{1}{w-z} = -\frac{1}{z(1-w/z)} = -\sum_0^\infty \frac{w^m}{z^{m+1}}$$

which in turn yields,

$$\frac{1}{2\pi\imath}\int_{C_1}\frac{f(w)}{w-z}dw = -\sum_0^\infty\left(\frac{1}{2\pi\imath}\int_{C_1} f(w)w^m dw\right) z^{-m-1} = -\sum_{-\infty}^{-1} c_n z^n \tag{5.16}$$

(by putting $n = -m-1$).

Finally, from lemma 5.4.3, it follows that

$$f(z) = \frac{1}{2\pi\imath}\int_{|w|=\rho_2}\frac{f(w)dw}{w-z} - \frac{1}{2\pi\imath}\int_{|w|=\rho_1}\frac{f(w)dw}{w-z} = \sum_{-\infty}^\infty c_n z^n,$$

where c_n are given by integrals as in the statement of the theorem. Since this is valid for all $r_1 < \rho_1 < \rho_2 < r_2$, it is valid in the annulus A. ♠

Remark 5.4.2

1. By shifting the origin at any other point $z = a$, we get statements similar to that of theorem 5.4.1, for annular regions around a.

2. Take the case $r_1 = 0$. Then $z = 0$ is an isolated singularity of f. This singularity is a pole of order, say, n iff the terms in the Laurent expansion vanish precisely below $-n$, i.e., $c_m = 0$ for all $m < -n$ and $c_{-n} \neq 0$. (Of course, if this n happens to be 0 then the singularity is removable.) This is the same as saying that the point $z = 0$ is an essential singularity of f iff $c_m \neq 0$ for infinitely many $m < 0$.

3. Put $f^+(z) := \dfrac{1}{2\pi\imath}\displaystyle\int_{C_2}\frac{f(w)dw}{w-z}$ and $f^-(z) := \dfrac{1}{2\pi\imath}\displaystyle\int_{C_1}\frac{f(w)dw}{w-z}$. Then we have, for $\rho_1 < |z| < \rho_2$,
$$f(z) = f^+(z) - f^-(z). \tag{5.17}$$
Clearly, f^+ is holomorphic in $|z| < \rho_2$ and f^- is holomorphic in $|z| > \rho_1$. Moreover, $f^-(z) \longrightarrow 0$ as $z \longrightarrow \infty$. Indeed, with these properties, using the uniqueness part of the above theorem or otherwise, we can see that f^+ and f^- are unique and are given by 5.15 and 5.16 respectively. They are called the *regular* and *principle* part of f respectively.

   Indeed, observe that f^+ is defined and holomorphic in $\mathbb{C} \setminus C_2$. Likewise, f^- is holomorphic in $\mathbb{C} \setminus C_1$. Also, for $|z| < \rho_1$, or $|z| > \rho_2$, it is easily seen that $f^+(z) = f^-(z)$. This property should not be confused to mean that f^+ and f^- are the same.

4. It is worth noting that often when we have to prove the *existence and uniqueness* of some mathematical property, the *uniqueness* comes handy in the proof of the *existence* part. The above theorem is one such instance.

5. The integral formula for c_n has tremendous theoretical importance. However, it is not so useful in actual computation of the Laurent series, for which, we should often depend on the geometric series expansion.

Example 5.4.1

1. Let us begin with a simple example first. Consider $f(z) = \dfrac{1}{1-z}$. Inside the disc $|z| < 1$ this has the geometric series representation,

$$f(z) = \sum_0^\infty z^n.$$

However, on the annulus $|z| > 1$, we have,

$$f(z) = -\sum_{-\infty}^{-1} z^n.$$

Let us get the Laurent series for f on $|z-2| > 1$. We have,

$$f(z) = -\frac{1}{z-1} = -\frac{1}{z-2+1} = -\frac{1}{z-2}\left(\frac{1}{1+\frac{1}{z-2}}\right) = \sum_{-\infty}^{-1}(-1)^n(z-2)^n.$$

2. Consider the function $f(z) = \dfrac{1}{1+z^2}$. This is holomorphic in $\mathbb{C} \setminus \{\imath, -\imath\}$. Thus, given any center c, there are, in general, three regions on which we can study the Laurent series. Of course, for some special c, two of the regions may coincide. For instance, for $c = 0$ we have $A_1 = \{z \ : \ |z| < 1\}$ and $A_2 = \{z \ : \ |z| > 1\}$. As usual, we get a power series in z on A_1 and a power series in $1/z$ on A_2, viz., $\sum_0^\infty (-1)^n z^{2n}$, and $\sum_{-\infty}^{-1}(-1)^{n-1}z^{2n}$ respectively. Now consider $c = 1+\imath$. Then we have to consider three different regions: $A_1 = \{z \ : \ |z-1-\imath| < 1\}$;
$A_2 = \{z \ : \ 1 < |z-1-\imath| < \sqrt{5}\}$ and
$A_3 = \{z \ : \ |z-1-\imath| > \sqrt{5}\}$.
The first thing we have to do is to express f in terms of partial fractions;

$$f(z) = -\frac{1}{2\imath(z+\imath)} + \frac{1}{2\imath(z-\imath)}. \tag{5.18}$$

In A_1, both the terms are holomorphic and hence we get power series in $(z-1-\imath)$.

$$\frac{1}{z-\imath} = \frac{1}{z-1-\imath+1} = \sum_0^\infty (-1)^n(z-1-\imath)^n; \tag{5.19}$$

$$\frac{1}{z+\imath} = \frac{1}{z-1-\imath+1+2\imath} = \frac{1}{1+2\imath}\sum_0^\infty (-1)^n\left(\frac{z-1-\imath}{1+2\imath}\right)^n. \tag{5.20}$$

Hence,

$$f(z) = \frac{1}{2\imath}\sum_{0}^{\infty}(-1)^n\left(1 - \frac{1}{(1+2\imath)^{n+1}}\right)(z-1-\imath)^n.$$

On the annulus A_2, (5.20) still holds. However, (5.19) is no more valid. Instead, we have,

$$\frac{1}{z-\imath} = \frac{1}{z-1-\imath}\sum_{0}^{\infty}(-1)^n\left(\frac{1}{z-1-\imath}\right)^n \tag{5.21}$$

Therefore,

$$f(z) = \frac{1}{2\imath}\left(\sum_{-\infty}^{-1}(-1)^{n-1}(z-1-\imath)^n + \sum_{0}^{\infty}(-1)^{n-1}\frac{(z-1-\imath)^n}{(1+2\imath)^{n+1}}\right).$$

We leave it to you to figure out the details for the region A_3.

3. Let $h(z) = \dfrac{1}{z(z-a)(z-b)}$, for some $0 < |a| < |b|$. Once again, with center 0, there are three regions to be discussed separately.

$$A_1 = \{z \ : \ 0 < |z| < |a|\}; \quad A_2 = \{z \ : \ |a| < |z| < |b|\}; \ A_3 = \{z \ : \ |z| > |b|\}.$$

By partial fractions, we have,

$$g(z) := ab(b-a)h(z) = \frac{b-a}{z} - \frac{b}{z-a} + \frac{a}{z-b}. \tag{5.22}$$

Consider the annulus A_1. We have,

$$\frac{b}{z-a} = -\frac{b}{a}\left(\frac{1}{1-z/a}\right) = -\frac{b}{a}\left(1 + \frac{z}{a} + \cdots + \frac{z^n}{a^n} + \cdots\right) \tag{5.23}$$

Likewise,

$$\frac{a}{z-b} = -\frac{a}{b}\left(1 + \frac{z}{b} + \cdots + \frac{z^n}{b^n} + \cdots\right) \tag{5.24}$$

Therefore,

$$\begin{aligned} h(z) &= \frac{b-a}{z} + \frac{b}{a}\left(1 + \frac{z}{a} + \cdots \frac{z^n}{a^n} + \cdots\right) - \frac{a}{b}\left(1 + \frac{z}{b} + \cdots + \frac{z^n}{b^n} + \cdots\right) \\ &= \frac{b-a}{z} + \sum_{n\geq 0}\left(\frac{b^{n+2} - a^{n+2}}{a^{n+1}b^{n+1}}\right)z^n. \end{aligned}$$

Next consider the annulus A_2. Here (5.23) is no longer valid. Instead, we have,

$$\frac{b}{z-a} = \frac{b}{z}\left(\frac{1}{1-a/z}\right) = \frac{b}{z}\left(1+\frac{a}{z}+\cdots+\frac{a^n}{z^n}+\cdots\right) \tag{5.25}$$

The third summand is the holomorphic part and hence (5.24) is still valid. Hence, we have,

$$\begin{aligned} h(z) &= \frac{b-a}{z} - \frac{b}{z}\left(1+\frac{a}{z}+\cdots+\frac{a^n}{z^n}+\cdots\right) - \frac{a}{b}\left(1+\frac{z}{b}+\cdots+\frac{z^n}{b^n}+\cdots\right) \\ &= -\frac{a}{z} - \frac{b}{a}\sum_{m\geq 2}\left(\frac{a}{z}\right)^m - \frac{a}{b}\sum_{n\geq 0}\left(\frac{z}{b}\right)^n. \end{aligned}$$

We leave it to the reader to write down the Laurent series for g in the annulus A_3, as an exercise.

Exercise 5.4

1. Obtain Laurent series expansion for the following functions in the respective annuli:
   (a) $f(z) = \dfrac{1}{1+z^2}$; $\{z \,:\, |z-1-\imath| > \sqrt{5}\}$.
   (b) $\dfrac{1}{z(z-a)(z-b)}$; $\{z \,:\, |z| > b\}$, $0 < a < b$.

2. Find the Laurent series representations of $\dfrac{1}{z(z-1)(z-2)}$ around $0, 1$, and 2 in all possible annuli.

3. Show that for all positive integers k, the Laurent series development of $\exp(z^{-k})$ in $\mathbb{C}\setminus\{0\}$ is
$$1+\frac{1}{z^k}+\frac{1}{2!z^{2k}}+\cdots+\frac{1}{n!z^{nk}}+\cdots.$$

4. Two Laurent series $\sum_{-\infty}^{\infty} a_n z^n$ and $\sum_{-\infty}^{\infty} b_n z^n$ converge to the same function in an open annulus. Show that $a_n = b_n$ for all integers.

5. What is the region of convergence of the Laurent series in each of the following cases:
   (a) $\displaystyle\sum_{-\infty}^{\infty}\frac{z^n}{2^{|n|}}$ (b) $\displaystyle\sum_{-\infty}^{\infty}\frac{z^n}{(|n|)!}$; (c) $\displaystyle\sum_{-\infty}^{\infty}\frac{(z-1)^{2n}}{n^2+1}$; (d) $\displaystyle\sum_{-\infty}^{\infty}\frac{(z-3)^{2n}}{(n^2+1)^n}$.

6. In each of the following cases, determine types of singularities and the principal parts:
   (a) $\dfrac{\sin z}{z^k}$, $k\in\mathbb{N}$; (b) $\cos(1/z)\sin(1/z)$.

5.5 Residues

Given a function f with an isolated singularity at a, and a circle C around a, putting $f = f^+ + f^-$, as in (5.17), we see that

$$\int_C f(z)dz = \int_C f^-(z)dz, \tag{5.26}$$

since the integral of the regular part vanishes. Also, since for any $n \geq 2$, the function z^{-n} has a primitive all over the circle $|z-a| = r$, we have

$$\int_{|z-a|=r} \frac{dz}{(z-a)^n} = 0. \tag{5.27}$$

Thus the only term which contributes to the integral (5.26) is the $1/z$-part of f which we shall now study.

Definition 5.5.1 Let f be a function on a domain Ω with the set of isolated singularities denoted by S. To each $a \in S$ consider an annulus region $0 < |z-a| < \delta$ contained in $\Omega \setminus S$. Let C be a circle around a, inside this annulus. We know that $\int_C \frac{1}{z-a} dz = 2\pi\imath$. (See 4.20.) Treating this as a normalizing factor, we define the numbers

$$\boxed{R_a(f) := \mathrm{Res}_a(f) = \frac{1}{2\pi\imath}\int_C f(z)\,dz.} \tag{5.28}$$

We call $R_a(f)$ the *residue* of f at $z = a$. Often, we may drop the function from the notation, when there is no confusion and use the notation R_a or Res_a.

Lemma 5.5.1 The residue of f at an isolated singularity $z = a$ is the unique number R_a such that the function

$$g(z) := f(z) - \frac{R_a}{z-a} \tag{5.29}$$

has a primitive in the whole of the annulus $0 < |z-a| < \delta$. Also, then R_a is equal to the coefficient of $(z-a)^{-1}$ in the Laurent expansion of f around a.

Proof: Let $f(z) = \sum_{-\infty}^{\infty} c_n (z-a)^n$, be the Laurent series representation of f in a disc around a. All powers of $(z-a)$ except $n = -1$ have primitives in the punctured disc, and term-by-term differentiation is valid. Hence, if g is as defined in (5.29), with $R_a = c_{-1}$, then g has a primitive. Also, since term-by-term integration is valid, it follows that

$$R_a(f) = \frac{1}{2\pi\imath}\int_C f(z)dz = \frac{1}{2\pi\imath}\int_C \frac{c_{-1}}{(z-a)}\,dz = c_{-1},$$

which also proves the uniqueness of $R_a(f)$ with the said property. ♠

The following theorem is of practical importance in computing the residues at poles, without writing down the Laurent expansion.

Theorem 5.5.1 *Let a be a pole of order n of f and let $g(z) = (z-a)^n f(z)$ with g holomorphic and $g(a) \neq 0$. Then the residue of f at a is given by*

$$R_a(f) = \frac{g^{(n-1)}(a)}{(n-1)!}.$$

Proof: We have, $f(z) = \dfrac{b_n}{(z-a)^n} + \cdots + \dfrac{b_1}{z-a} + g_n(z)$, and hence after multiplying by $(z-a)^n$ we obtain,

$$g(z) = b_n + \cdots + b_1(z-a)^{n-1} + g_n(z)(z-a)^n.$$

Differentiate $(n-1)-$times and put $z = a$ to get the result. ♠

Example 5.5.1

1. Let $f(z) = e^z/(z^2-1), z \neq \pm 1$. Then $z = \pm 1$ are simple poles of f. To compute the residue at $z = 1$, we write $g(z) = e^z/(z+1)$ and find $g(1) = e/2$. Therefore $\mathrm{Res}_1 = e/2$. Similarly $R_{-1} = -e^{-1}/2$.

2. Let $f(z) = (\sinh z)/z^3 := (e^z - e^{-z})/2z^3$. Clearly $z = 0$ is a pole. What is the order of this pole? Caution is needed in this type of examples. For, sinh has a zero of order 1 at 0. Hence it follows that the order of the pole of f at 0 is 2. Therefore the residue is given by the value of $((\sinh z)/z)'$ at $z = 0$. This can be computed using L'Hospital's rule (Exercise 5.1.1), as follows:

$$\begin{aligned} R_o = \lim_{z \longrightarrow 0} ((\sinh z)/z)' &= \lim_{z \longrightarrow 0} \frac{z \cosh z - \sinh z}{z^2} \\ &= \lim_{z \longrightarrow 0} \frac{\cosh z - z(\sinh z) - \cosh z}{2z} \\ &= \frac{1}{2} \lim_{z \longrightarrow 0} (-z \cosh z - \sinh z) = 0. \end{aligned}$$

   Alternatively, the Taylor' expansion can be employed, whenever the method above becomes cumbersome, for instance, when the order of the pole is high. In this example, we know that $\sinh z = z + \dfrac{z^3}{3!} + \dfrac{z^5}{5!} + \cdots$. Therefore, it follows immediately that the $(1/z)$-term is missing from the Laurent expansion of $\dfrac{\sinh z}{z^3}$. Hence, $R_0 = 0$. Use this method now to show that $\mathrm{Res}_{z=0} \dfrac{\sinh z}{z^4} = \dfrac{1}{6}$.

3. Consider the case when $f(z) = g(z)p(z)$ where g is given by a Laurent series and p is a polynomial:
$$g(z) = \sum_{-\infty}^{\infty} a_n z^n; \qquad p(z) = \sum_{0}^{m} \alpha_k z^k.$$
Then the residue of f at 0 is given by
$$R_0(f) = a_{-1}\alpha_0 + \cdots + a_{-k-1}\alpha_k + \cdots + a_{-m-1}\alpha_m.$$
For example, if $g(z) = e^{1/z}$ then, the residue of f at 0 is :
$$\frac{\alpha_m}{(m+1)!} + \cdots + \frac{\alpha_1}{2!} + \alpha_0.$$

4. Let f have a zero (or a pole) of order n (respectively $-n$) at a point $z = a$. Then $R_a(f'/f) = n$ (resp,$-n$.): To see this, let m be the algebraic order of f at a. Then we know that $f(z) = (z-a)^m g(z)$ for a holomorphic function g with $g(a) \neq 0$. Now differentiate and divide by f to see that
$$\frac{f'(z)}{f(z)} = \frac{m}{z-a} + \frac{g'(z)}{g(z)}.$$
Since $g(a) \neq 0$, g'/g is holomorphic at a. Hence, f'/f has a simple pole at a with $R_a(f'/f) = m$. This m is equal to $\pm n$ according as a is zero or a pole of f. This is often referred to as the logarithmic residue of f at $z = a$, because $\frac{d}{dz}(\ln f(z)) = \frac{f'(z)}{f(z)}$.

5. Let f have a simple pole at z_0 and g be holomorphic. Then $R_{z_0}(fg) = g(z_0)R_{z_0}(f)$. To see this, write
$$f(z) = \frac{b_{-1}}{z - z_0} + \sum_{0}^{\infty} b_j (z-z_0)^j; \quad g(z) = \sum_{0}^{\infty} c_j (z-z_0)^j,$$
valid in a neighborhood of z_0. Clearly, the Laurent series for fg which is the Cauchy product of these two, has the coefficient of $(z-z_0)^{-1}$ equal to $c_0 b_{-1}$.

Exercise 5.5 Compute the residues at all the singular points of the following functions:
(a) $\tan z$ (b) $\cot z$ (c) $\sec z$ (d) $e^{1/z}$ (e) $\dfrac{z}{\sin z}$
(f) $\dfrac{4}{1-z}$ (g) $\dfrac{e^z}{(z+\pi\imath)^3}$ (h) $\dfrac{z}{z^3-1}$ (i) $\dfrac{5}{(z^2-1)^2}$ (j) $\dfrac{e^{\imath z^2}}{(1-\cos z)^2}$.

5.6 Winding Number

Cauchy's formula (4.32) is a direct consequence of (4.31). In view of the discussion in the previous section about residues, the integral on the RHS of (4.31) assumes more importance. Indeed in order to bring out the true strength of (4.31), we need to understand the integral on the right side of (4.31) thoroughly. In this section, we take up this task, cook up a name and symbol for this integral, see some immediate applications, and finally give a geometric recipe to compute this integral.

Lemma 5.6.1 Let γ be a closed contour not passing through a given point z_0. Then the integral $w = \displaystyle\int_\gamma \frac{dz}{z - z_0}$ is an integer multiple of $2\pi i$.

Proof: Enough to prove that $e^w = 1$. Define

$$\alpha(t) := \int_a^t \frac{\gamma'(s)}{\gamma(s) - z_0}\, ds; \quad g(t) = e^{-\alpha(t)}(\gamma(t) - z_0); a \leq t \leq b.$$

Since γ is continuous and differentiable except at finitely many points, so is g. Moreover, wherever g is differentiable, we have

$$\begin{aligned} g'(t) &= -e^{-\alpha(t)}\alpha'(t)(\gamma(t) - z_0) + e^{-\alpha(t)}\gamma'(t) \\ &= e^{-\alpha(t)}(-\gamma'(t) + \gamma'(t)) \;=\; 0. \end{aligned}$$

Therefore, $g(t) = g(a) = \gamma(a) - z_0,$ for all $t \in [a, b]$ and hence,

$$e^{\alpha(t)} = \frac{\gamma(t) - z_0}{\gamma(a) - z_0},$$

for all $t \in [a, b]$. Since $\gamma(a) = \gamma(b)$, it now follows that $e^w = e^{\alpha(b)} = e^{\alpha(a)} = 1.$ ♠

Definition 5.6.1 Let γ be a closed contour not passing through a point z_0. Put

$$\int_\gamma \frac{dz}{z - z_0} \;=\; 2\pi\imath m.$$

Then the number m is called the *winding number of the closed contour γ around the point z_0* and is denoted by $\eta(\gamma, z_0)$. Thus

$$\boxed{\eta(\gamma, z_0) := \frac{1}{2\pi\imath}\int_\gamma \frac{dz}{z - z_0}.} \tag{5.30}$$

Remark 5.6.1 In order to understand the concept of winding number let us examine it a little closely.

1. Take $z_0 = 0$ and γ to be any circle around 0. Then we have seen that
$$\int_\gamma \frac{dz}{z} = 2\pi\imath.$$
In other words, $\eta(\gamma, 0) = 1$. So we can say that γ *winds* around 0 exactly once and this coincides with our geometric intuition.

2. Now let γ be any simple closed contour contained in the interior of an open disc in the upper half plane. Since $1/z$ is holomorphic in that disc, by corollary 4.2.2 or otherwise (it has a primitive), it follows that $\displaystyle\int_\gamma \frac{dz}{z} = 0$. That means $\eta(\gamma, 0) = 0$. Hence in this case, we see that the winding number is zero which again conforms with our geometric understanding.

3. More generally, if γ is contained in a disc, then for all points z outside this disc, we have $\eta(\gamma, z) = 0$. This is a simple consequence of Cauchy's theorem 4.3.3 for discs or by simply observing that $1/(z - a)$ has a primitive on the disc. Once again this conforms with our general understanding that a contour inside a disc does not go around any point z outside the disc.

4. Let us now consider the curve $\gamma(t) = e^{2\pi\imath nt}$, defined on the interval $[0, 1]$ for some integer n. This curve traces the unit circle $n-$times in the counter clockwise direction. This tallies with the computation of
$$\int_\gamma \frac{dz}{z} = 2\pi\imath n.$$

5. As a direct consequence of (4.9) and (4.10), we have
$$\boxed{\boxed{\eta(\gamma_1.\gamma_2, z) = \eta(\gamma_1, z) + \eta(\gamma_2, z); \qquad \eta(\gamma^{-1}, z) = -\eta(\gamma, z).}} \tag{5.31}$$

6. By Prop. 4.1.2, it follows that $z \mapsto \eta(\gamma, z)$ is a continuous function on $\mathbb{C} \setminus Im(\gamma)$. Being an integer valued continuous function, it must be locally constant. Therefore, by 1.6.1, it is a constant function on each connected subset of $\mathbb{C} \setminus Im(\gamma)$.

7. Enclosing γ in a large circle C and taking a point z_0 outside C, it follows from (3) that $\eta(\gamma, z_0) = 0$. Hence, it follows that $\eta(\gamma, z) = 0$ for all points z in the unbounded component of $\mathbb{C} \setminus Im(\gamma)$. Moreover, any unbounded component should intersect $\mathbb{C} \setminus B_r(0)$ for large r and hence there is only one unbounded component of $\mathbb{C} \setminus Im(\gamma)$. For this reason, we can define $\eta(\gamma, \infty) = 0$, so that $\eta(\gamma, -)$ gets extended continuously on $\widehat{\mathbb{C}} \setminus Im(\gamma)$.

8. The following special case of theorem 4.5.1 is of utmost importance: Assume that for some component Ω_1 of $\mathbb{C}\setminus Im(\gamma)$, we have, $\eta(\gamma, w) = 1,\ \forall\, w \in \Omega_1$. (See Fig.26.) Then on Ω_1, f itself is represented by

$$f(w) = \frac{1}{2\pi\imath}\int_\gamma \frac{f(z)\,dz}{z-w}, \quad \forall\, w \in \Omega_1. \tag{5.32}$$

In particular, the holomorphic function f is completely determined on this region by the value of f on γ.

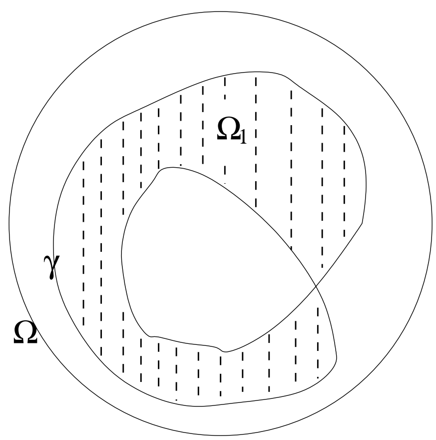


Fig. 26

Example 5.6.1 Let us find the value of

$$\int_{|z|=1} \frac{e^{az}}{z-a}\,dz.$$

Observe that e^{az} is holomorphic on the entire plane. The integral makes sense for all points a such that $|a| \neq 1$. For points $|a| < 1$, the curve γ defining the unit circle has the property $\eta(\gamma, a) = 1$ and for those points a such that $|a| > 1$ we have $\eta(\gamma, a) = 0$. Hence, by (5.32), the given integral is equal to $2\pi\imath e^{a^2}$ for $|a| < 1$ and 0 for $|a| > 1$.

Example 5.6.2 As a simple minded application of theorem 5.6.1, let us prove the non existence of certain roots. Assume that Ω is a domain which contains a closed contour $\gamma : [a, b] \longrightarrow \mathbb{C}$, such that $\eta(\gamma, 0)$ is odd. Then we claim that there does not exist any holomorphic function $g : \Omega \longrightarrow \mathbb{C}$ such that $g^2(z) = z, z \in \Omega$. Let us assume on the contrary. Then by differentiating, we get, $2g(z)g'(z) = 1, z \in \Omega$. Now,

$$\eta(g\circ\gamma, 0) = \frac{1}{2\pi\imath}\int_{g\circ\gamma}\frac{dw}{w} = \frac{1}{2\pi\imath}\int_a^b \frac{g'(\gamma(t))\gamma'(t)}{g(\gamma(t))}dt = \frac{1}{4\pi\imath}\int_a^b \frac{\gamma'(t)}{\gamma(t)}dt = \frac{\eta(\gamma,0)}{2}.$$

This means that $\eta(\gamma, 0)$ is even which is absurd. Similar statements will be true for other roots also, viz., we do not have a well defined n^{th} root of $z - z_0$ in any domain that contains a closed contour γ such that $\eta(\gamma, z_0)$ is not divisible by n. [Later, we shall see that existence of such a closed contour implies the existence of a closed contour γ with $\eta(\gamma, z_0) = 1$.]

Example 5.6.3 Let us now consider the function $f(z) = 1 - z^2$ and study the question when and where there is a holomorphic single valued branch g of the square root of f i.e., $g^2 = f$. Observe that $z = \pm 1$ are the zeros of f and hence if these points are included in the region then there would be trouble: By differentiating the identity $g^2 = f$, we obtain $2g(z)g'(z) = f'(z) = -2z$. This is impossible since, at $z = \pm 1$, the L.H.S.$= 0$ and R.H.S. $= \mp 2$. So the region on which we expect to find g should not contain ± 1.

Next, assume that Ω contains a small circle C around 1, say, contained in a punctured disc $\Delta' := B_\epsilon(1) \setminus \{1\}, 0 < \epsilon < 1$. Restricting our attention to Δ', observe that there is a holomorphic branch of the square root of $1 + z$ say h defined all over $B_\epsilon(1)$. Clearly $h(z) \neq 0$ here and hence $\phi = g/h$ will then be a holomorphic function on $\Delta' \cap \Omega$ such that $\phi^2 = 1 - z$. This contradicts our observation in the example 5.6.2.

By symmetry, we conclude that Ω cannot contain any circle which encloses only one of the points $-1, 1$.

Finally, suppose that both ± 1 are in the same connected component of $\mathbb{C} \setminus \Omega$. Then for all closed contours γ in Ω, both ± 1 will be in the same connected component of $\mathbb{C} \setminus Im(\gamma)$ and hence $\eta(\gamma, 1) = \eta(\gamma, -1)$. For instance, take $\Omega = \mathbb{C} \setminus [-1, 1]$. Then for any circle C with center 0 and radius > 1, $\eta(C, 1) = \eta(C, -1) = 1$.

We shall now see that the square root of f exists. Consider the flt $T(z) = \dfrac{1 - z}{1 + z}$. This maps $\mathbb{C} \setminus [-1, 1]$ onto $\mathbb{C} \setminus \{x \in \mathbb{R} \ : \ x \leq 0\}$, on which we can choose a well defined branch of the square root function. This amounts to say that we have a holomorphic function $h : \mathbb{C} \setminus [-1, 1] \longrightarrow \mathbb{C}$ such that $h(z)^2 = \dfrac{1 - z}{1 + z}$. Now consider $g(z) = h(z)(1 + z)$. Then $g(z)^2 = f(z)$, as required.

In fact, $\Omega(= \mathbb{C} \setminus [-1, 1])$ happens to be a maximal domain on which $1 - z^2$ has a well defined square root. This follows from our earlier observation that any such domain on which g exists cannot contain a circle which encloses only one of the two points $-1, 1$.

Finally, observe that, in place of $[-1, 1]$, if we had any arc joining -1 and 1, the image of such an arc under T would be an arc from 0 to ∞ and hence on the complement of it, square-root would still exist. Also, the above discussion holds verbatim to the function $(z - a)(z - b)$ for any $a \neq b \in \mathbb{C}$. You can also modify this argument to construct other roots. Now it is time for you to take a look at the exercise 1 below.

Remark 5.6.2 In view of remark (5.6.1.8), we shall give a sufficient condition for a contour to have winding number ± 1 around a point. This condition is quite a practical one in the sense that it is easy to verify it in many concrete situations. In the statement of the lemma below, we have simply assumed that $z_0 = 0$. Of course, this does not diminish the generality of the result, as we can always perform a translation and choose the origin to be any given point. The result is important from application as well as theoretical point of view. *However, you may skip learning the proof of this for the time being and come to it later.*

Lemma 5.6.2 Let γ be a contour not passing through 0. Let z_1, z_2 be two distinct points on γ and let L be a directed line through 0 so that z_1 and z_2 are on the opposite sides of L. Denote the two portion of the curve γ between z_1 and z_2, by ω_1 and ω_2 so that we have $\gamma = \omega_1.\omega_2$. Assume further that ω_1 does not meet the *negative ray* of L and ω_2 does not meet the *positive ray* of L. Then

$$\frac{1}{2\pi\imath}\int_\gamma \frac{dz}{z} = \eta(\gamma, 0) = \pm 1.$$

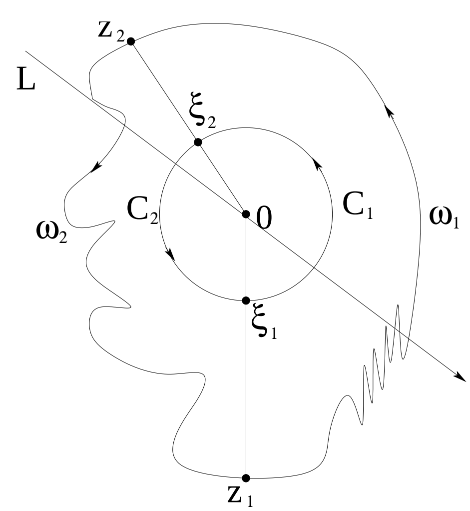


Fig. 27

Proof: Let C be a circle around 0, not meeting γ and let $\xi_1,\ \xi_2$ be the points on C lying on the line segments $[0, z_1]$ and $[0, z_2]$ respectively. If C_1 and C_2 are the portion of the circle traced counter clockwise from ξ_1 to ξ_2 and from ξ_2 to ξ_1 respectively, it follows that C_1 does not meet the negative ray of L and C_2 does not meet the positive ray of L. Let

$$\tau_1 = [\xi_1, z_1].\omega_1.[z_2, \xi_2].(C_1)^{-1}; \quad \tau_2 = (C_2)^{-1}.[\xi_2, z_2].\omega_2.[z_1, \xi_1].$$

Then it follows that τ_i are closed contours and

$$\eta(\tau_1, 0) + \eta(\tau_2, 0) = -\eta(C, 0) + \eta(\gamma, 0).$$

On the other hand, since τ_1 does not meet the negative ray of L, it follows that 0 is in the unbounded component of $\mathbb{C} \setminus Im(\tau_1)$ and hence as observed in remark 5.6.1.7, this implies that $\eta(\tau_1, 0) = 0$. For similar reason $\eta(\tau_2, 0) = 0$. Therefore,

$$\eta(\gamma, 0) = \eta(C, 0) = 1.$$

If we had taken the other orientation on γ, we would have got $\eta(\gamma, 0) = -1$. This completes the proof of the lemma. ♠

As an immediate application we have:

Theorem 5.6.1 *Let $\Omega \subset \mathbb{C}$ be a bounded convex region with a smooth boundary C oriented counter clockwise. Then*

$$\eta(C; a) = \begin{cases} 1, & a \in \Omega \\ 0, & a \in \mathbb{C} \setminus \bar{\Omega}. \end{cases}$$

In particular, this is true for any disc and any rectangle.

Proof: First consider $a \in \Omega$. Any line L through a cuts C into two parts. Now for any two points z_1, z_2 on C lying on opposite sides of L, the hypothesis of the above lemma is easily verified. This gives the first part.

Now appeal to the fact that $\mathbb{C} \setminus C$ has two components, one of which is Ω and the other is $\mathbb{C} \setminus \bar{\Omega}$. (Compare Ex. 1.6.3.) By remark 5.6.1.7, the second part follows. ♠

Exercise 5.6

1. Prove or disprove that $f(z) = 1 - z^2$ has a well defined logarithm in $\mathbb{C} \setminus [-1, 1]$.

2. Let γ be a closed contour, p be a point not on γ, and L be an infinite ray beginning at p and intersecting γ in exactly n points at each of these points 'crossing' it over to the other side. Show that $\eta(\gamma, a) \equiv n \bmod(2)$. Give a recipe to determine the actual value of $\eta(\gamma, a)$ from these considerations.

3. Evaluate $\displaystyle\int_\gamma (e^z - e^{-z}) z^{-4}\, dz$, where γ is one of the closed contours drawn below:

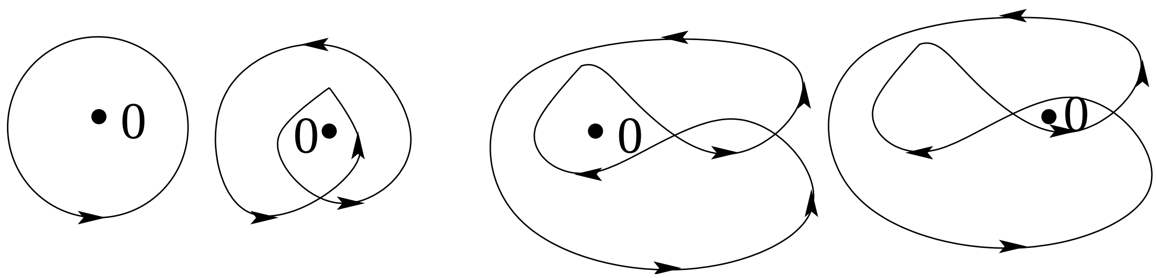


Fig. 28

5.7 The Argument Principle

We have seen that integration along suitably chosen contours detect residue at a point. Combined with Cauchy's theorem, and the concept of winding number, this can be made into an effective tool. Here is the first step:

Theorem 5.7.1 Residue Theorem I-Version: *Let Ω be a convex domain in $\mathbb{C}$, S be a finite subset of Ω and let $\Omega' = \Omega \setminus S$. Let γ be a closed contour in Ω'.*

Then for any holomorphic function f on Ω', we have

$$\frac{1}{2\pi\imath}\int_\gamma f dz = \sum_{a\in S} \eta(\gamma, a)R_a(f). \tag{5.33}$$

Proof: Let $S = \{a_1, \ldots, a_m\}$ and

$$g_j(z) := \sum_{k=-1}^{-\infty} c_{k,j}(z - a_j)^k$$

be the principal part of f at a_j. Then $f - \sum_{j=1}^m g_j$ is holomorphic on Ω. Each g_j is holomorphic in $\Omega \setminus \{a_j\}$ and hence, term-by-term integration is valid. This gives,

$$\int_\gamma g_j(z)dz = \int_\gamma \frac{c_{-1,j}}{z - a_j}dz = 2\pi\imath\eta(\gamma, a_j)c_{-1,j} = 2\pi\imath R_{a_j}(f)\eta(\gamma, a_j).$$

On the other hand, by Cauchy's Theorem 4.3.3, since $f - \sum_j g_j$ is holomorphic in Ω, $\displaystyle\int_\gamma \left(f - \sum_j g_j\right) dz = 0$ which means that

$$\int_\gamma f(z)dz = \sum_{j=1}^k \int_\gamma g_j dz = 2\pi\imath \sum_{j=1}^k \eta(\gamma, a_j)R_{a_j}(f).$$

This completes the proof. ♠

Remark 5.7.1 We can replace the hypothesis 'S be a finite subset of Ω' in the above theorem to the hypothesis 'S is a discrete subset of Ω' as follows. Observe that this part is purely topological in nature.

Lemma 5.7.1 Let Ω be a convex open set and γ be a closed contour in it. Suppose S is a discrete subset of Ω which does not intersect $Im(\gamma)$. Then $\eta(\gamma, a) = 0$ for all but finitely many $a \in S$.

Proof: If not, suppose $\{a_n\}$ is a sequence of points in A with $\eta(\gamma, a_n) \neq 0$. We may assume that this sequence lies in a bounded component of $\mathbb{C}\setminus Im(\gamma)$ (see Remark 5.6.1.7.) By Bolzano-Weierstrass property, we may pass on to a subsequence which converges and denote its limit by a. Since A is a discrete subset Ω, this implies that $a \in \mathbb{C} \setminus \Omega$. Since Ω is convex, it follows that $\eta(\gamma, a) = 0$ by Cauchy's theorem 4.3.3. Since $\eta(\gamma, -)$ is locally constant, it follows that $\eta(\gamma, a_n) = 0$ for $n >> 1$, which is a contradiction. ♠

Lemma 5.7.2 Given any compact convex subset K and a closed subset T of a convex domain Ω, such that $K \cap T = \emptyset$, there exists a convex open set $U(K) \subset \Omega$ such that the closure $\overline{U(K)}$ is compact, $K \subset U(K)$ and $U(K) \cap T = \emptyset$.

Proof: Let δ be the distance between K and T. Then $\delta > 0$. Take $U(K) = \{z \in \Omega \ : \ d(z, K) < \delta\}$ and verify all the claims. ♠

Theorem 5.7.2 Residue Theorem II-Version: *Let Ω be a convex domain in $\mathbb{C}$, S be a discrete subset of Ω and let $\Omega' = \Omega \setminus S$. Let γ be a closed contour in Ω'. Then for any holomorphic function f on Ω', we have*

$$\frac{1}{2\pi\imath}\int_\gamma f dz = \left(\sum_{a \in S} \eta(\gamma, a) R_a(f)\right). \tag{5.34}$$

Proof: Let K be set of all finite sums

$$\sum_j t_j z_j; \quad \sum_j t_j = 1, \quad 0 \le t_j \le 1, \quad z_k \in Im(\gamma)$$

i.e., the convex hull of the compact set $Im(\gamma)$. Since $Im(\gamma)$ is bounded, it follows that K is bounded. Clearly K is a closed subset also. Since S is a discrete subset, $K \cap S$ is finite and $T = S \setminus K \cap S$ is a closed subset. (See remarks 4.3.2.) By the above lemma, we get a convex nbd $U(K)$ of K such that $T \cap U(K) = \emptyset$. Therefore $U(K) \cap S \subset K \cap S$ is finite. Now we can apply the I-version of the residue theorem with Ω replaced by $U(K)$. ♠

The next step is to combine this result with remark 4.

Theorem 5.7.3 Logarithmic Residue Theorem: *Let g be a holomorphic function and f be a meromorphic function which is not identically 0 in a convex domain Ω. Let $a_1, \ldots, a_k, \ldots$ and $b_1, \ldots, b_l, \ldots$ be a listing of the zeros and poles of f, with each zero and pole being repeated as many times as its order in the listing. Then for any closed contour γ in Ω which does not pass through any of the a_j, b_k's, we have,*

$$\frac{1}{2\pi\imath}\int_\gamma \frac{g(z)f'(z)}{f(z)}dz = \sum_j \eta(\gamma, a_j)g(a_j) - \sum_k \eta(\gamma, b_k)g(b_k). \tag{5.35}$$

Proof: Let $\mathcal{I}$ denote the set of zeros and poles of f. Then we know that $\mathcal{I}$ is an isolated subset of Ω and points of $\mathcal{I}$ are precisely the set of poles of $f'g/f$ each of which is simple. We apply theorem 5.7.2 above taking gf'/f in place of f. Moreover, for each $c \in \mathcal{I}$ we have, $R_c(f'/f)$ is equal to the algebraic multiplicity of f at c. Now from the example 5.5.1.5, it follows that

$$R_{a_j}(gf'/f) = n_j g(a_j); \quad R_{b_k}(gf'/f) = -n_k g(b_k)$$

where n_j and n_k are the orders of f at a_j and b_k respectively. Therefore, from theorem 5.7.2, we have,

$$\begin{aligned}\frac{1}{2\pi\imath}\int_\gamma \frac{f'(z)}{f(z)}g(z)\,dz &= \sum_{c\in\mathcal{I}} \eta(\gamma, c)R_c(gf'/f) \\ &= \sum_j \eta(\gamma, a_j)g(a_j) - \sum_k \eta(\gamma, b_k)g(b_k).\end{aligned}$$

♠

Remark 5.7.2

1. An interesting consequence of this is when f is holomorphic in a disc D and not vanishing on ∂D. Then

$$\boxed{\frac{1}{2\pi\imath}\int_{\partial D} \frac{f'(z)}{f(z)}dz = \text{ total number of zeros of } f \text{ inside } D.} \tag{5.36}$$

2. The above result has the following interesting interpretation: Let $\Gamma(t) = f \circ \gamma(t)$, where $\gamma(t)$, $0 \le t \le 1$, is a parameterization of ∂D, traced counter clockwise. Then Γ is the image of γ under f. By definition, the winding number of Γ around 0 is given by,

$$\begin{aligned}\eta(\Gamma, 0) &= \frac{1}{2\pi\imath}\int_\Gamma \frac{dw}{w} = \frac{1}{2\pi\imath}\int_0^1 \frac{(f\circ\gamma)'(t)}{f(\gamma(t))}dt \\ &= \frac{1}{2\pi\imath}\int_0^1 \frac{f'(\gamma(t))\gamma'(t)}{f(\gamma(t))}dt = \frac{1}{2\pi\imath}\int_\gamma \frac{f'(z)}{f(z)}dz.\end{aligned}$$

In the figure below, there are three simple roots of $f(z)$ inside the disc.

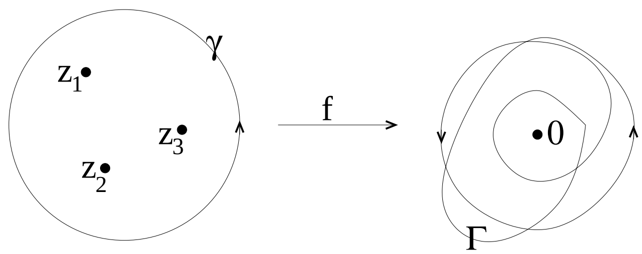


Fig. 29

3. Thus *f maps γ onto a curve Γ which winds around 0 as many times as there are zeros of f inside γ.* This is called the **argument principle.** When you have a pole, the orientation gets reversed as in the case of $z \mapsto 1/z$ and hence, the curve is winding around 0 in the clockwise direction. The theorem is a slight generalization of this result. In a latter chapter, (see Exercise 7.4,) we shall have a version which is even more general and call it Generalized Argument Principle. A useful go-in-between needs to be paid proper attention too:

Theorem 5.7.4 *Let f be a meromorphic function which is not identically* 0 *on a closed disc D, and having no zeros and poles on ∂D. Then the winding number of the contour $\Gamma = f \circ \gamma$, where $\gamma = \partial D$ is equal to the number of zeros of f minus the number of poles of f in the interior of D, each counted with its multiplicity.*

Proof: All that you have to do is to recall that $\eta(\Gamma, 0)$ is defined by the formula

$$\eta(\Gamma, 0) = \frac{1}{2\pi\imath}\int_\Gamma \frac{dw}{w} = \frac{1}{2\pi\imath}\int_\gamma \frac{f'(z)}{f(z)}dz.$$

Now, remember $\eta(\gamma, a) = 1$ for all $a \in D$ and apply the logarithmic residue theorem with $g(z) = 1$. ♠

Example 5.7.1 Let $p = a^2$ be an isolated singularity of f and let $g(z) = 2zf(z^2)$. Then clearly both $\pm a$ are isolated singularities of g. Let us see what is the relation between the residues $R_p(f)$ and $R_{\pm a}(g)$.

First consider the case when $a \neq 0$. Let γ denote a positively oriented small closed contour around a consisting of four portions of hyperbola as in the example 3.6.2 and the accompanying figure. It follows that

$$2\pi\imath R_a(g) = \int_\gamma g(z)dz = \int_0^1 2\gamma(t)f((\gamma(t)^2)\gamma'(t)dt. \tag{5.37}$$

Let us denote the curve $t \mapsto (\gamma(t))^2$ by γ_1 say. Then γ_1 is nothing but the boundary of an axial rectangle around p. Therefore, $\eta(\gamma_1, p) = 1$. Hence, (5.37) yields,

$$2\pi\imath R_a(g) = \int_0^1 f(\gamma_1(t))\gamma_1'(t)dt = \int_{\gamma_1} f(z)dz = 2\pi\imath R_p(f). \tag{5.38}$$

Clearly, this is the case at $-a$ as well.

Next consider the case $a = 0$. Here we can choose γ to be a small circle with center at 0. It then follows that γ_1 is a circle around 0 traced twice, i.e., $\eta(\gamma_1, 0) = 2$. Therefore (5.37) now yields,

$$2\pi\imath R_0(g) = \int_{\gamma_1} f(z)dz = 4\pi\imath R_0(f). \tag{5.39}$$

Therefore, we have,

$$R_{\pm a}(2zf(z^2)) = \begin{cases} R_{a^2}(f), & a \neq 0 \\ 2R_{a^2}(f) & a = 0. \end{cases}$$

In either case, the residue $R_{a^2}(f)$ is equal to the sum of the residues $R_{\pm a}(zf(z^2))$. In conclusion, more generally, if S is the set of all isolated singularities of $zf(z^2)$ then $S^2 = \{z^2 \ : \ z \in S\}$ is the set of all isloated singularities of f. Further, in case one of them is finite, we have,

$$\sum_{b\in S^2} R_b(f(z)) = \sum_{a\in S} R(zf(z)). \tag{5.40}$$

Exercise 5.7 Use residues to compute the integrals $\displaystyle\int_C f(z)dz$, where C denotes the unit circle traced in the counter clockwise sense and f is given by:

(a) $e^{1/z}$ (b) $\dfrac{e^z}{z}$ (c) $\dfrac{z}{2z-1}$ (d) $\dfrac{\sin \pi z}{z^6}$ (e) $\cot z$

(f) $\dfrac{\tanh z}{e^z \sin z}$ (g) $\dfrac{\sin^6 z}{(z-\pi/6)^3}$ (h) $\dfrac{z}{(z^2+4z+1)^2}$. (i) $z \sin\left(\dfrac{1}{z}\right)$ (j) $\dfrac{1}{(1-\cos z)^6}$.

5.8 Miscellaneous Exercises to Ch. 5

1. Let f be an entire function such that $\displaystyle\lim_{z\to\infty} \frac{f(z)}{z} = 0$. Show that f is a constant.

2. Let f be an entire function such that $|f(z)| \leq A|z|$ for some real constant A. Show that $f(z) = az$ for some $|a| < A$.

3. Let f be an entire function such that $|f(z)| \leq 1+|z|^{1/2}$. Show that f is a constant. Can you replace 1/2 by some other positive real number?

4. Let f be an entire function. Suppose $\lim\limits_{z\to\infty} \dfrac{\Re(f(z))}{z} = 0$. Show that f is a constant.

5. Let A be an isolated set in $\mathbb{C}$. Show that a bounded holomorphic function $f : \mathbb{C} \setminus A \to \mathbb{C}$ is a constant.

6. Consider the triangle OPQ where $O = (0,0), P = (\pi, 0), Q = (0,1)$. Find the maximum of the following functions on this triangle. (i) $|z^2+1|$; (ii) e^z.

7. Let $\{a_n\}$ be a sequence of complex numbers such that $\lim_{n\to\infty} |a_n|^{1/n} = L,\ 0 < L < \infty$. Consider the function defined by the series $f(z) = \sum_n a_{2n} z^n$. Suppose there exists an entire function $g : \mathbb{C} \to \mathbb{C}$ such that $f(z)g(z) = 1, |z| < 1,\ g(5) = 0$ and $g(z) \neq 0$, for $|z| < 5$. Determine the value of L.

Chapter 6

Application to Evaluation of Definite Real Integrals

In this chapter, we shall demonstrate the usefulness of the complex integration theory in computing definite real integrals. This should not surprise you since after all, complex integration is nothing but two real integrals which make up its real and imaginary parts. Thus given a real integral to be evaluated if we are successful in associating a complex integration and also evaluate it, then all that we have to do is to take real (or the imaginary) part of the complex integral so obtained. However, this itself does not seem to be always possible. Moreover, as we think about it, we perceive several obstacles in this approach. For instance, the complex integration theory is always about integration over closed paths whereas, a real definite integral is always over an interval, finite or infinite. So, by adding suitable curves, we somehow form a closed curve, on which the complex integration is performed and then we would like either to get rid of the value of the integration on the additional paths that we have introduced or we look for other sources and methods to evaluate them. The entire process is called *'the method of complexes'.* Each problem calls for a certain amount of ingenuity. Thus we see that the method has its limitations and as Ahlfors puts it "*— but even complete mastery does not guarantee success.*" However, when it works it works like magic.

All said and done, it should be pointed out that the method of complexes is not always the best. This point will be illustrated in an exercise (8.3.2) in chapter 8, by evaluating integrals $\displaystyle\int_0^\infty \frac{\sin x}{x}\,dx$ in an amusing way.

Let us learn this chapter through discussion of some standard examples. On the way, we shall keep recording down the wisdom that we accumulate in the form of theorems.

A careful study of the examples discussed should enable you to write down the proof of these theorems by yourself. We implore you to try it and compare with the solution given at the end.

In the following, we use the abbreviation 'RT' for the residue theorem.

6.1 Trigonometric Integrals

Let us show that
$$\int_0^{2\pi} \frac{d\theta}{1+a\sin\theta} = \frac{2\pi}{\sqrt{1-a^2}}, \ -1 < a < 1.$$

Observe that for $a = 0$, there is nothing to prove. So let us assume that $a \neq 0$. We want to convert the integrand into a function of a complex variable and then set $z = e^{\imath\theta}$, $0 \leq \theta \leq 2\pi$, so that the integral is over the unit circle C. Since, $z = e^{\imath\theta} = \cos\theta + \imath\sin\theta$, we have, $\sin\theta = (z - z^{-1})/2\imath$, and $dz = \imath e^{\imath\theta}d\theta$, i.e., $d\theta = dz/\imath z$. Therefore,
$$I = \int_C \frac{dz}{\imath z(1 + a(z - z^{-1})/2\imath))} = \int_C \frac{2dz}{az^2 + 2\imath z - a} = \frac{2}{a}\int_C \frac{dz}{(z - z_1)(z - z_2)},$$
where, z_1, z_2 are the two roots of the polynomial $z^2 + 2\imath z/a - 1$. Note that $z_1 = (-1 + \sqrt{1-a^2})\imath/a$, $z_2 = (-1 - \sqrt{1-a^2})\imath/a$. It is easily seen that $|z_2| > 1$. Since $z_1z_2 = -1$, it follows that $|z_1| < 1$. Therefore on the unit circle C the integrand has no singularities and the only singularity inside the circle is a simple pole at $z = z_1$. The residue at this point is given by
$$R_{z_1} = \frac{2}{a(z_1 - z_2)} = \frac{-\imath}{\sqrt{1-a^2}}.$$
Hence by the RT we have:
$$I = 2\pi\imath R_{z_1} = \frac{2\pi}{\sqrt{1-a^2}}.$$

A proof of the following theorem can be extracted easily from what we have seen so far:

Theorem 6.1.1 Trigonometric integrals : *Let $\phi(x,y) = p(x,y)/q(x,y)$ be a rational function in two variables such that $q(x,y) \neq 0$ on the unit circle. Then*
$$I_\phi := \int_0^{2\pi} \phi(\cos\theta, \sin\theta)d\theta = 2\pi\left(\sum_{z\in\mathbb{D}} R_z(\tilde{\phi})\right),$$
where, $\tilde{\phi}(z) = \dfrac{1}{z}\phi\left(\dfrac{z+z^{-1}}{2}, \dfrac{z-z^{-1}}{2\imath}\right)$.

Exercise 6.1

1. Write down an explicit proof of theorem 6.1.1.

2. Use the complex method to prove the following, where $a, b \in \mathbb{R}$:

(i) $\displaystyle\int_0^{2\pi} \frac{d\theta}{5 + 4\cos\theta} = \frac{2\pi}{3}.$ $\qquad(ii)$ $\displaystyle\int_0^{2\pi} \frac{d\theta}{1 + a\cos\theta} = \frac{2\pi}{\sqrt{1-a^2}},\ |a| < 1.$

(iii) $\displaystyle\int_{-\pi}^{\pi} \frac{d\theta}{1 + \sin^2\theta} = \sqrt{2}\pi.$ $\qquad(iv)$ $\displaystyle\int_0^{2\pi} \frac{d\theta}{(a + \cos\theta)^2} = \frac{2a\pi}{(a^2-1)^{3/2}},\ a > 1.$

(v) $\displaystyle\int_0^{2\pi} \frac{d\theta}{(2 + cos\theta)^2} = \frac{4\pi}{\sqrt[3/2]{3}}.$ $\qquad(vi)$ $\displaystyle\int_0^{2\pi} \frac{d\theta}{a + b\cos\theta} = \frac{2\pi}{\sqrt{a^2-b^2}},\ |b| < a.$

6.2 Improper Integrals

We shall begin with a brief introduction to the theory of improper integrals. Chiefly there are two types of them. One type arises due to the infiniteness of the interval on which the integration is being taken. The other type arises due to the fact that the integrand is not defined (shoots to infinity) at one or both end point of the interval.

Definition 6.2.1 When $\displaystyle\int_a^b f(x)dx$ is defined for all $b > a$ we define

$$\int_a^\infty f(x)dx := \lim_{b \longrightarrow \infty} \int_a^b f(x)dx, \tag{6.1}$$

if this limit exists. Similarly we define

$$\int_{-\infty}^b f(x)dx := \lim_{a \longrightarrow -\infty} \int_a^b f(x)dx, \tag{6.2}$$

if this limit exists. Also, we define

$$\int_{-\infty}^\infty f(x)dx := \int_0^\infty f(x)dx + \int_{-\infty}^0 f(x)dx, \tag{6.3}$$

provided both the integrals on the right exist.

Recall the Cauchy's criterion for the limit. It follows that the limit (6.1) exists iff *given* $\epsilon > 0$, there exists $R > 0$ such that for all $d > c > R$ we have,

$$\left|\int_c^d f(x)dx\right| < \epsilon. \tag{6.4}$$

In many practical situations the following theorem and statements which can be easily derived out of it come handy in ensuring the existence of the improper integral of this type.

Theorem 6.2.1 of Improper Integrals : *Suppose f is a continuous function defined on $[0, \infty)$ and there exists $\alpha > 1$ such that $x^\alpha f(x)$ is bounded. Then $\int_0^\infty f(x)\,dx$ exists.*

Proof: Observe that the improperness is only at ∞. There, the convergence follows by comparing with the integral $\int_a^\infty \frac{1}{x^\alpha} dx$. ♠

However, the condition in the above theorem is not always necessary. For instance, the function $f(x) = \frac{\sin x}{x}$ does not satisfy this condition. Nevertheless $\int_0^\infty \frac{\sin x}{x}\,dx$ exists as will be seen soon.

Observe that there is yet another legitimate way of taking limits in (6.3), i.e., to take the limit of $\int_{-a}^{a} f(x)dx$, as $a \longrightarrow \infty$. However, this limit, even if it exists, is, in general, not equal to the improper integral defined in 6.3, above. This is called the *Cauchy's Principal Value* of the improper integral and is denoted by,

$$PV\left(\int_{-\infty}^{\infty} f(x)dx\right) := \lim_{a \longrightarrow \infty} \int_{-a}^{a} f(x)dx. \tag{6.5}$$

As an example consider $f(x) = x$. Then the Cauchy's PV exists but the improper integral does not. However, if the improper integral exists, then it is also equal to its principle value. This observation is going to play a very important role in the following application. [Indeed, if we let ourselves consider the line integrals, only in the sense of Cauchy principle value (at improper points), then we have wider applicability of Cauchy's theory. We shall see this through some exercises at the end of the section]. As typical example of this, consider the case when the integrand f is an even function i.e., $f(-x) = f(x)$, for all x, then we see that if the Cauchy's PV exists then the improper integral exists and is equal to PV.

Example 6.2.1 Let us consider the problem of evaluating

$$I = \int_0^\infty \frac{2x^2 - 1}{x^4 + 5x^2 + 4} dx$$

Denoting the integrand by f, we first observe that f is an even function and hence

$$I = \frac{1}{2} \int_{-\infty}^{\infty} f(x)dx$$

which in turn is equal to its PV. Thus we can hope to compute this by first evaluating

$$I_R = \int_{-R}^{R} f(x)dx$$

and then taking the limit as $R \longrightarrow \infty$. First we extend the rational function into a function of a complex variable so that the given function is its restriction to the real axis. This is easy here, viz., consider $f(z)$. Next we join the two end points R and $-R$ by an arc in the upper-half space, (no harm if you choose the lower half-space). What could be a better way than to do it with the semi-circle! So let C_R denote the semi-circle running from R to $-R$ in the upper-half space. Let γ_R denote the closed contour obtained by tracing the line segment from $-R$ to R and then tracing C_R. We shall compute

$$J_R = \int_{\gamma_R} f(z)dz$$

for large R using residue computation. When the number of singular points of the integrand is finite, J_R is a constant for all large R. This is the crux of the matter. We then hope that in the limit, the integral on the unwanted portions tends to zero, so that $\lim_{R\longrightarrow\infty} J_R$ itself is equal to I.

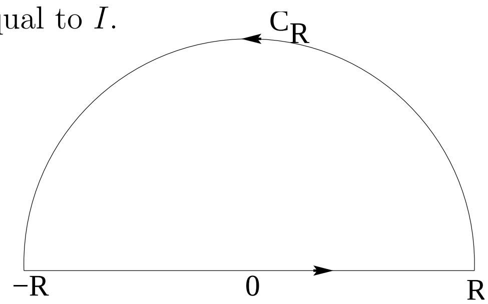


Fig. 30

The first step is precisely where we use the residue theorem. The zeros of the denominator $q(z) = z^4 + 5z^2 + 4$ are $z = \pm\imath, \;\pm 2\imath$ and *luckily* they do not lie on the real axis.(This is important.) They are also different from the roots of the numerator. Also, for $R > 2$, two of them lie inside γ_R. (We do not care about those in the lower half-space.) Therefore by the RT, we have, $J_R = 2\pi\imath(R_\imath + R_{2\imath})$. The residue computation easily shows that $J_R = \pi/2$.

Observe that $f(z) = p(z)/q(z)$, where $|p(z)| = |2z^2 - 1| \;\leq\; 2R^2 + 1$, and similarly $|q(z)| = |(z^2+1)(z^2+4)| \;\geq\; (R^2-1)(R^2-4)$. Therefore

$$|f(z)| \;\leq\; \frac{2R^2+1}{(R^2-1)(R^2-4)} \;=:\; M_R.$$

This is another lucky break that we have got. Note that M_R is a rational function of R of degree -2. For, now we see that

$$|\int_{C_R} f(z)dz| \leq M_R \int_{C_R} |dz| = M_R R\pi.$$

Since M_R is of degree -2, it follows that $M_R R\pi \longrightarrow 0$ as $R \longrightarrow \infty$. Thus, we have successfully shown that the limit of $\int_{C_R} f(z)dz$ vanishes at infinity. To sum up, we have,

$$I = \frac{1}{2}\int_{-\infty}^{\infty} f(x)dx = \frac{1}{2}\lim_{R\longrightarrow\infty}\int_{-R}^{R} f(x)dx = \frac{1}{2}\lim_{R\longrightarrow\infty} J_R = \frac{\pi}{4}.$$

Indeed, we have seen enough to write down a proof of the following theorem.

Theorem 6.2.2 *Let f be a meromorphic function in a domain containing the closure of the upper half plane $\boldsymbol{H}$, with finitely many poles in $\boldsymbol{H}$. Suppose $\lim_{z\to\infty} zf(z) = 0$ (where z is taken inside the closed upper half plane $\Im(z) \geq 0$.) Then*

$$PV\left(\int_{-\infty}^{\infty} f(x)\,dx\right) = 2\pi\imath \sum_{w\in\boldsymbol{H}} R_w(f).$$

In particular, this is so, if $f(z) = p(z)/q(z)$ is a rational function of degree ≤ -2 and $q(r) \neq 0$ for any $r \in \mathbb{R}$.

Example 6.2.2 Let us consider another example which is indeed covered by the above theorem, yet is not exactly similar to the earlier example:

$$\int_{-\infty}^{\infty} f(x)dx$$

where $f(x) = (\cos 3x)(x^2+1)^{-2}$.

Except that now the integrand is a rational function of a trigonometric quantity and the variable x, this does not seem to cause any trouble as compared to the example above. For we can consider $F(z) = e^{3\imath z}(z^2+1)^{-2}$ to go with and later take only the real part of whatever we get. The denominator has poles at $z = \pm\imath$ which are double poles but that need not cause any concern. When $R > 1$ the contour γ_R encloses $z = \imath$ and we find the residue at this point of the integrand, and see that $J_R = 2\pi/e^3$. Yes, the bound that we can find for the integrand now has different nature! Putting $z = x + \imath y$ we know that $|e^{3\imath z}| = |e^{-3y}|$. Therefore,

$$|f(z)| = \left|\frac{e^{3\imath z}}{(z^2+1)^2}\right| \leq \left|\frac{e^{-3y}}{(R^2-1)^2}\right|.$$

Since, e^{-3y} remains bounded by 1 for all $y > 0$ we are done. Thus it follows that the given integral is equal to $2\pi/e^3$.

Exercise 6.2

1. Show that (a) $\displaystyle\int_{-\infty}^{\infty} \frac{x^2}{1+x^4}dx = \frac{\pi}{\sqrt{2}}$; (b) $\displaystyle\int_{-\infty}^{\infty} \frac{x \sin ax \, dx}{x^4+4} = \frac{\pi}{2}e^{-a}(\sin a)$.
2. Evaluate (a) $\displaystyle\int_0^{\infty} \frac{\cos x}{x^2+a^2}\, dx,\ a > 0;$ (b) $\displaystyle\int_0^{\infty} \frac{x^2 dx}{x^4+x^2+1}$.
3. Write down an explicit proof for theorem 6.2.2.

6.3 Jordan's Inequality

In the previous section, we had several lucky breaks. The next step is going to get us into some real trouble. Consider the problem of evaluating the Cauchy's Principal Value of

$$I = \int_{-\infty}^{\infty} f(x)dx, \quad \text{where,} \quad f(x) = (x \sin\, x)/(x^2+2x+2).$$

Writing $f(x) = g(x)\sin x$ and taking $F(z) = g(z)e^{\imath z}$, we see that, for $z = x$, we see that $f(x) = \Im(F(x))$. Also, write, $g(z) = z/(z^2+2z+2) = z/(z-z_1)(z-z_2)$ where, $z_1 = \imath - 1$ and $z_2 = -\imath - 1$, to see that $|g(z)| \leq R/(R-\sqrt{2})^2 =: M_R,\ R > 2$, say. And of course, this implies that $\displaystyle\int_{C_R} F(z)dz$ is bounded by $\pi R M_R$, which does not tend to zero as $R \longrightarrow \infty$. Hence, this is of no use! Thus, we are now forced to consider the following stronger estimate:

Lemma 6.3.1 Jordan's Inequality

$$J := \int_0^{\pi} e^{-R\sin\theta}d\theta \ < \ \pi/R, \quad R \ > \ 0.$$

Proof: Draw the graph of $y = \sin\theta$ and $y = 2\theta/\pi$. Conclude that $\sin\theta \ > \ 2\theta/\pi$, for $0 \ < \theta < \pi/2$. Hence obtain the inequality,

$$e^{-R\sin\theta} \ < \ e^{-2R\theta/\pi}$$

Use this to obtain,

$$J := 2\int_0^{\pi/2} e^{-R\sin\theta}d\theta \ < \ 2\int_0^{\pi/2} e^{-2R\theta/\pi}d\theta = \frac{2\pi(1-e^{-R})}{2R} \ < \ \pi/R,\ \ R \ > \ 0.$$

♠

Let us now use this in the computation of the integral I above. We have

$$\begin{aligned} \left|\int_{C_R} F(z)dz\right| &= \left|\int_0^\pi g(Re^{\imath\theta})e^{\imath Re^{\imath\theta}}\imath Re^{\imath\theta}d\theta\right| \\ &< M_R R\int_0^\pi e^{-R\sin\theta}d\theta < M_R\pi. \end{aligned} \tag{6.6}$$

Since $M_R\pi \longrightarrow 0$ as $R \longrightarrow \infty$, we get

$$\Im(\lim_{R\longrightarrow\infty} J_R) = I,$$

as required. We leave the calculation of the residue to the reader.
[Answer: $\dfrac{\pi}{e}(\cos 1 + \sin 1)$.]

We now have enough ideas to prove:

Theorem 6.3.1 *Let f be a holomorphic function in $\mathbb{C}$ except possibly at finitely many singularities none of which is on the real line. Suppose that $\lim_{z\to\infty} f(z) = 0$. Then for any non zero real a,*

$$PV\left(\int_{-\infty}^{\infty} f(x)e^{\imath ax}\,dx\right) = \pm 2\pi\imath \sum_{\pm w\in \boldsymbol{H}} R_w[f(z)e^{\imath az}],$$

where, the sign $\pm$ has to be chosen (in both places), according as a is positive or negative.

Remark 6.3.1 We should also add that the conditions of the theorem are met if f is a rational function of degree ≤ -1 having no real poles. See the Misc. Exercises for a different method.

Exercise 6.3

1. Write down an explicit proof of theorem 6.3.1.

2. Let $a > 0$ and $\Im w \neq 0$. Then, show that

$$\int_{-\infty}^{\infty} \frac{e^{\imath ax}}{x-w}\,dx = \begin{cases} 2\pi\imath e^{\imath aw}, & \text{if } \quad w \in \boldsymbol{H} \\ 0, & \text{if } -w \in \boldsymbol{H}. \end{cases}$$

3. Use the above exercise to get the following formulae due to Laplace:

$$\int_0^\infty \frac{b\cos ax}{x^2+b^2}dx = \int_0^\infty \frac{x\sin ax}{x^2+b^2}dx = \frac{1}{2}\pi e^{-ab}, \quad a, b > 0.$$

[From this, Cauchy deduced the following:

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$$

by substituting $b = 0$. What kind of justification was needed in this argument?]

4. Show that $\displaystyle\int_{-\infty}^{\infty} \frac{x \sin x + \cos x}{x^2 + 1} = \frac{2\pi}{e}$.

6.4 Bypassing a Pole

Here we shall attempt to evaluate $\displaystyle\int_0^\infty \frac{\sin x}{x} dx$.

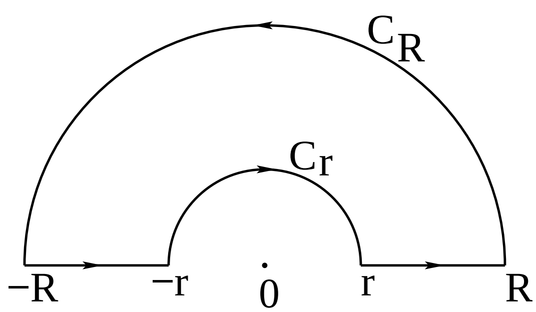


Fig. 31

First of all observe that $\dfrac{\sin x}{x}$ is an even function and hence,

$$\int_0^\infty \frac{\sin x}{x}\, dx = \frac{1}{2} PV \left(\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx \right).$$

The associated complex function $F(z) = e^{\imath z}/z$ has a singularity on the x-axis and that is going to cause trouble if we try to proceed the way we did so far. Common sense tells us that, since 0 is the point at which we are facing trouble, we should simply avoid this point by going around it via a small semi-circle around 0 in the upper half-plane. Thus consider the closed contour $\gamma_{r,R}$ as shown in the Fig. 31.

The idea is to

(i) compute $I(r, R) := \displaystyle\int_{\gamma_{r,R}} F(z) dz$,

(ii) take the limit as $r \longrightarrow 0$ and $R \longrightarrow \infty$,

(iii) show that the integral on the larger circular portion tends to zero [see (6.6)] and finally,

(iv) take the imaginary part of the result.

Since $F(z)$ is holomorphic inside of $\gamma_{r,R}$, it follows that the integral is zero for all $R > r > 0$. Thus the value of the given integral is equal to half of the imaginary part of $-\lim_{r\longrightarrow 0}\int_{C_r} F(z)dz$. Thus, it remains to compute this limit. Since 0 is a simple pole of $F(z)$ we can write $zF(z) = g(z)$ with $g(0) \neq 0$. Again using Taylor's theorem, write $g(z) = g(0) + zg_1(z)$ where g_1 is holomorphic in a neighborhood of 0. It follows that $F(z) = g(0)/z + g_1(z)$. Therefore,

$$\int_{C_r} F(z)dz = g(0)\int_0^{\pi} \imath d\theta \; + \int_{C_r} g_1(z)dz = g(0)\pi\imath + (G_1(r) - G_1(-r))$$

where , G_1 is a primitive of g_1 in a disc around 0. By continuity of G_1, the last term tends to zero as $r \longrightarrow 0$. So it remains only to compute $g(0)$ which is nothing but the residue of $e^{\imath z}/z$ at $z = 0$. Thus we obtain

$$\int_0^{\infty} \frac{\sin x}{x} = \frac{\pi}{2}. \tag{6.7}$$

Moral: When the contour of integration enclosed a pole, $2\pi\imath$ times the residue was added to the value of the integral. However, when the contour passes through a pole only a portion of the above quantity was added. At least, in case the contour had a straight line segment at such a pole, this portion was exactly half. A natural question that arises is: *Is it so when the contour smooth but not necessarily straight?; is it so if it has a singularity?* See the exercises at the end of the chapter.)

Combining our observations in this section with that in the previous one, we obtain:

Theorem 6.4.1 *Let f be meromorphic in $\mathbb{C}$ with finitely many singularities in the closure of the upper half space. Suppose* $\lim_{z\longrightarrow\infty} f(z) = 0$. *Then for any $a > 0$ we have,*

$$PV\left(\int_{-\infty}^{\infty} f(x)e^{\imath ax}dx\right) = \pi\imath\left(2\sum_{w\in \boldsymbol{H}} R_w[f(z)e^{\imath az}] + \sum_{w\in\mathbb{R}} R_w[f(z)e^{\imath az}]\right).$$

Exercise 6.4

1. Find the value of $\displaystyle\int_{-\infty}^{\infty} \frac{\sin x}{x(x^2-1)}\,dx$.

2. Find the value of $\displaystyle\int_{-\infty}^{\infty} \frac{\cos x}{x^2-1}\,dx$.

3. Write down a an explicit proof of theorem 6.4.1.

6.5 Inverse Laplace Transforms

The purpose of this section is to establish the integral formula for the inverse Laplace transform, as an application of residue calculus.

First, we shall briefly recall the Method of Laplace Transform in solving an initial value problem:

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y + a_n = 0;\ y^{(j)}(0) = b_j, j = 0, \ldots, n-1. \tag{6.8}$$

Given a function $f : [0, \infty) \to \mathbb{R}$ the **Laplace transform** $\mathcal{L}(f) = F$ is defined by

$$F(s) = \int_0^\infty e^{-st} f(t)\, dt \tag{6.9}$$

provided the improper integral on the RHS exists. This is a mild but non trivial condition and let us consider only such functions f here and denote the class of such functions by $\mathcal{A}$. (See Ex. 6.5. 2 below.)

Two basic properties of $\mathcal{L}$ are of interest to us:

(i) **Linearity** Given $f, g \in \mathcal{A}, a, b \in \mathbb{R}$, we have

$$\mathcal{L}(af + bg) = a\mathcal{L}(f) + b\mathcal{L}(g). \tag{6.10}$$

(ii) **Derivative Formula:** For $f \in \mathcal{A}$ which is continuously differentiable n times,

$$\mathcal{L}(f^{(j)})(s) = s^j \mathcal{L}(f)(s) - s^{j-1} f(0) - s^{j-2} f'(0) - \cdots - f^{(j-1)}(0), \quad 1 \le j \le n. \tag{6.11}$$

Suppose now that $y = f(t)$ is the solution of (6.8) and let $Y = Y(s) = \mathcal{L}(y)(s)$. By the linearity of $\mathcal{L}$ it follows that

$$\mathcal{L}(y^{(n)}) + a_1 \mathcal{L}(y^{n-1)}) + \cdots + a_{n-1}\mathcal{L}(y) + a_n \mathcal{L}(1) = 0.$$

Observe that the constant function $t \mapsto 1$ has its Laplace transform $\mathcal{L}(1)(s) = \frac{1}{s}$. Multiply (6.11) by a_j (with $a_0 = 1$) and sum over j to get an expression of the form

$$Y(s) = \frac{P(s)}{Q(s)} \tag{6.12}$$

where P and Q are polynomials of degree n and $n+1$. Indeed,

$$Q(s) = s^{n+1} + a_1 s^n + \cdots + a_n s + a_n;$$

$$P(s) = b_0 s^n + (b_1 + a_1 b_0) s^{n-1} + \cdots + (b_{n-1} + a_1 b_{n-2} \cdots + a_{n-1} b_0) s + a_n.$$

The success of the Method of Laplace transform hinges upon whether we can now determine y from the expression for Y. The task has become simpler since we should only know the answer to:

Q. Given a rationals function $\frac{P}{Q}$ of degree -1, determine y such that $\mathcal{L}(y) = \frac{P}{Q}$.

Following the popular practice, we simply call y as the Inverse Laplace transform of Y and write $\mathcal{L}^{-1}(Y) = y$ if $\mathcal{L}(y) = Y$. (Strictly speaking, this need a lot of justification.)

Theoretically, every rational function can be expressed in terms of partial fractions. Using the linearity of the $\mathcal{L}$, it is then enough to know the answer to the above question for rational functions of the form $Q(s) = \frac{1}{(s-a)^n}$. The table below gives $\mathcal{L}(f)$ for some simple elementary functions.

$f(t)$	$\mathcal{L}(f)(s)$	$f(t)$	$\mathcal{L}(f)(s)$
k, a constant	$\frac{k}{s}$	$t^n,\ n > 0$ an integer	$\frac{n!}{s^{n+1}}$
e^{at}	$\frac{1}{s-a}$	$t^n e^{at}$	$\frac{n!}{(s-a)^{n+1}}$
$\cos at$	$\frac{s}{s^2+a^2}$	$\sin at$	$\frac{a}{s^2+a^2}$
$\cosh at$	$\frac{s}{s^2-a^2}$	$\sinh at$	$\frac{a}{s^2-a^2}$

Example 6.5.1

Consider the initial value problem:

$$y'' + y' + y = 0, \quad y(0) = 1, \quad y'(0) = 0. \tag{6.13}$$

Denoting $Y = \mathcal{L}(y)$ we have

$$\mathcal{L}(y') = sY - y(0) = sY - 1; \mathcal{L}(y'') = s^2Y - sy(0) - s^2y'(0) = s^2Y - s.$$

Therefore we get

$$s^2Y + sY + Y = s + 1.$$

Therefore $Y = \frac{s+1}{s^2+s+1}$. We now express Y into partial fractions:

$$Y(s) = \frac{s+1}{s^2+s+1} = \frac{1}{\omega - \omega^2}\left(\frac{1+\omega}{s-\omega} - \frac{1+\omega^2}{s-\omega^2}\right).$$

Therefore,

$$y(t) = \mathcal{L}^{-1}(Y) == \frac{1}{\omega - \omega^2}[(1+\omega)e^{\omega t} - (1+\omega^2)e^{\omega^2}t] = e^{t/2}\left[\cos\frac{\sqrt{3}}{2}t + \frac{1}{\sqrt{3}}\sin\frac{\sqrt{3}}{2}t\right].$$

Remark 6.5.1 Observe that in the above table $\mathcal{L}^{-1}(F)$ is equal to the residue of $e^{tz}F(z)$ at the pole of F. Therefore, it follows that for any meromorphic function with finitely many poles, $z_1, z_2, \ldots, z_k$ (by expressing it in partial fractions) we have,

$$\mathcal{L}^{-1}(F)(t) = \sum_{j=1}^{k} \mathrm{Re}s_{z_j}[e^{tz}F(z)]. \tag{6.14}$$

Thus given a meromorphic function F with finitely many poles, we have an integral formula for the RHS of (6.14): We have to merely take a large loop γ which goes around all the poles of F exactly once, integrate $e^{zt}F(z)$ on it and divide by $2\pi\imath$. Let $\gamma = \gamma_R$ be the loop consisting of the vertical line segment L_R between $\alpha - \imath R$ and $\alpha + \imath R]$ and the circular arc $C_R := (\alpha + R\imath, R - \alpha, \alpha - \imath R)$ of radius R and center α, where $\alpha, R >> 0$ so that all the poles of F are inside γ_R. Let M_R be an upper bound for $|F(z)|$ on the circular part.

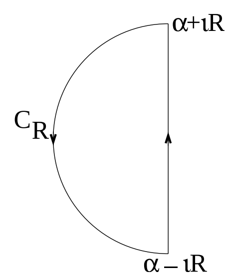


Fig. 32

Then using Jordan's inequality we can see that

$$\left|\int_{C_R} e^{zt}F(z)dz\right| \le e^{\alpha t}M_R R\int_{\pi/2}^{3\pi/2} e^{Rt\cos\theta}d\theta < e^{\alpha t}M_R\pi/t.$$

We now put the additional condition on F that $M_R \to 0$ as $R \to \infty$. (As seen before, this condition holds for rational functions of degree -1.) Then

$$f(t) = \frac{1}{2\pi}\lim_{R\to\infty}\int_{-R}^{R} e^{t(\alpha+s\imath)}F(\alpha+\imath s)ds = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{t(\alpha+\imath s)}F(\alpha+\imath s)ds \tag{6.15}$$

for all sufficiently large $\alpha > 0$. This gives the integral formula for the inverse Laplace operator. So, given a function F, we define inverse Laplace transform $\mathcal{L}^{-1}(F)$ by the formula

$$\mathcal{L}^{-1}(F) = \frac{1}{2\pi}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{t(\alpha+\imath s)}F(\alpha+\imath s)ds \tag{6.16}$$

whenever the RHS of (6.16) exists. What we have seen is that in the special case of $F = \frac{P}{Q}$ where P and Q are polynomials such that $deg\, Q - deqP > 0$, $\mathcal{L}(\mathcal{L}^{-1}(F)) = F$.

The all important question that 'if $\mathcal{L}(f_1) = \mathcal{L}(f_2)$ then is it true $f_1 = f_2$?' has to be determined separately. The answer is known to be 'almost yes' in the sense that $f_1 - f_2 \equiv 0$ except on an isolated set.

Exercise 6.5

1. Solve the following initial value problems using the Laplace transforms:

   (a) $y''' + y'' - 6y' = 0;\ y(0) = 0 = y''(0);\ y'(0) = 1.$

   (b) $y'' - 3y' + 2y = 3t + e^{3t};\ y(0) = 1, y'(0) = -1.$

2. Let $f : [0, \infty) \to real$ be a piecewise continuous function. Suppose there exist constant k, L such that $|f(t)| \leq Le^{kt}, t \geq 0$. Then show that $\mathcal{L}(f)$ exists.

6.6 Branch Cuts or Keyhole Integrals

Consider the problem of evaluating the integral

$$I = \int_0^\infty \frac{x^{-\alpha}}{x+1} dx, \quad 0 < \alpha < 1.$$

This integral is important in the theory of Gamma functions $\Gamma(a) = \displaystyle\int_0^\infty x^{a-1}e^{-x}dx$. Observe that the integral converges because in $[0, 1]$, we can compare it with $\displaystyle\int_0^1 x^{-\alpha}dx$, whereas, in $[1, \infty)$, we can compare it with $\displaystyle\int_1^\infty x^{-\alpha-1}dx$. The problem that we face here is that the corresponding complex function $f(z) = z^{-\alpha}$ does not have any single valued branch in any neighborhood of 0. So, an idea is to cut the plane along the positive real axis, take a well defined branch of $z^{-\alpha}$, perform the integration along a contour as shown in the figure below and then let the cuts in the circles tend to zero. The crux of the matter lies in the following observation:

Let $f(z)$ be a branch of z^α in $\mathbb{C} \setminus \{x : x \geq 0\}$. Suppose for any $x_0 > 0$, the limit of $f(z)$ as $z \longrightarrow x_0$ through upper-half plane is equal to $x_0^{-\alpha}$. Then the limit of $f(z)$ as $z \longrightarrow x_0$ through lower-half plane is equal to $x^{-\alpha}e^{-2\pi\imath\alpha}$.

This easily follows from the periodic property of the exponential. Now, let us choose a branch $f(z)$ of $z^{-\alpha}$, say, for which $f(-1) = e^{-\pi\imath\alpha}$ and integrate $g(z) = \dfrac{f(z)}{z+1}$ along the closed contour as shown in the figure. The figure justifies the name 'key-hole integral'.

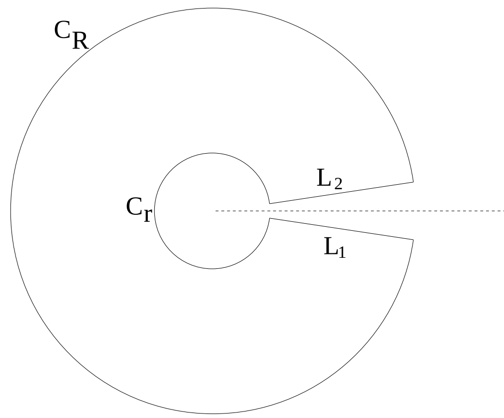


Fig. 33

When the radius r of the inner circle is smaller than 1 and radius R of the outer one is bigger that 1, this contour goes around the only singularity of $g(z)$ exactly once, in the counter clockwise sense. Hence,

$$\int_{\gamma} \frac{f(z)}{z+1}\, dz = 2\pi \imath e^{-\pi\imath\alpha} \tag{6.17}$$

We now let the two segments L_1, L_2 approach the interval $[r, R]$. This is valid, since in a neighborhood of $[r, R]$, there exist continuous extensions f_1 and f_2 of g_1 and g_2 where g_1 and g_2 are restrictions of g to upper half plane and lower half plane respectively. The RHS of the above equation remains unaffected where as on the LHS, we get,

$$\int_r^R \frac{x^{-\alpha}}{x+1}\, dx + \int_{|z=R|} \frac{f(z)}{z+1}\, dz - \int_r^R \frac{x^{-\alpha}e^{-2\pi\imath\alpha}}{x+1}\, dx - \int_{|z|=r} \frac{f(z)}{z+1}\, dz = 2\pi\imath e^{-\pi\imath\alpha}.$$

Now, we let $r \longrightarrow 0$ and $R \longrightarrow \infty$. It is easily checked that the two integrals on the two circles are respectively bounded by the quantities $2\pi R^{1-\alpha}/(R+1)$ and $2\pi r^{1-\alpha}/(r+1)$. Hence the limits of these integrals are both 0. Therefore,

$$(1 - e^{-2\pi\imath\alpha}) \int_0^\infty \frac{x^{-\alpha}}{x+1}\, dx = 2\pi\imath e^{-\pi\imath\alpha}.$$

Hence,

$$\int_0^\infty \frac{x^{-\alpha}}{x+1}\, dx = \frac{\pi}{\sin \pi\alpha}, \quad 0 < \alpha < 1.$$

There are different ways of carrying out the branch cut. See for example the book by Churchill and Brown, for one such. We shall cut out all this and describe yet another method here, which is most elegant.

Theorem 6.6.1 *Let ϕ be a meromorphic function on $\mathbb{C}$ having finitely many poles none of which belongs to $[0, \infty)$. Let $a \in \mathbb{C} \setminus \mathbb{Z}$ be such that $\lim_{z\to 0} z^a\phi(z) = 0 = \lim_{z\to\infty} z^a\phi(z)$. Then the following integral exists and*

$$I_a := \int_0^\infty x^{a-1}\phi(x)\,dx = \frac{2\pi\imath}{1 - e^{2\pi\imath a}} \sum_{w\in\mathbb{C}} Res_w(z^{a-1}\phi(z)). \tag{6.18}$$

Proof: First substitute $x = t^2$ and see that

$$I_a = \int_0^\infty x^{a-1}\phi(x)dx = 2\int_0^\infty t^{2a-1}\phi(t^2)dt. \tag{6.19}$$

Next choose a branch $g(z)$ of z^{2a-1} in $-\pi/2 < arg z < 3\pi/2$. Observe that $g(-x) = (-1)^{2a-1}g(x) = -e^{2\pi\imath a}g(x)$, for $x > 0$. Hence,

$$\begin{aligned}
\int_{-\infty}^{\infty} z^{2a-1}\phi(z^2)dz &= \int_0^\infty g(x)\phi(x^2)dx + \int_{-\infty}^0 g(x)\phi(x^2)dx \\
&= \int_0^\infty g(x)\phi(x^2)dx - \int_0^\infty e^{2\pi\imath a}g(x)\phi(x^2)dx \\
&= (1 - e^{2\pi\imath a})\int_0^\infty z^{2a-1}\phi(z^2)dz
\end{aligned}$$

Therefore,

$$I_a = \frac{2}{1 - e^{2\pi\imath a}} \int_{-\infty}^{\infty} z^{2a-1}\phi(z^2)dz = \frac{4\pi\imath}{1 - e^{2\pi\imath a}} \sum_{z\in\boldsymbol{H}} Res_z(z^{2a-1}\phi(z^2)). \tag{6.20}$$

If we set $f(z) = z^{a-1}\phi(z)$ then $zf(z^2) = z^{2a-1}\phi(z^2)$. We have seen in example 5.7.1 that the sum of the residues of $f(z)$ and that of $zf(z^2)$ are the same. Since, in this case, f has no poles on the real line, it follows that the sum of the residues of $zf(z^2)$ in $\boldsymbol{H}$ is equal to half the sum of the residues of f. The formula 6.18 follows. ♠

It may be noted that the assignment $a \mapsto I_a$ is called *Mellin's transform* corresponding to ϕ. Coming back to the special case when $\phi(z) = \dfrac{1}{z+1}$, we have $Res_{-1}\dfrac{z^{a-1}}{z+1} = (-1)^{a-1} = -e^{\pi\imath a}$. Hence,

$$\int_0^\infty \frac{x^{(a-1)}}{x+1} = \frac{\pi}{\sin \pi a}, \quad 0 < a < 1. \tag{6.21}$$

Observe that the condition that a is not an integer is crucial for the non existence of the branch of z^{a-1} throughout a neighborhood of 0. On the other hand, that is what guarantees the existence of the integral.

Exercise 6.6

1. Show that $\displaystyle\int_0^\infty \frac{x^{a-1}}{x+e^{\imath\theta}}dx = \frac{\pi}{\sin \pi a}e^{\imath(a-1)\theta}$, $0 < a < 1$, $-\pi < \theta < \pi$. Now, substitute $x = t^n$ and $a = m/n$, for integers $0 < m < n$ to obtain the formula:
$$\int_0^\infty \frac{x^{m-1}}{x^n + e^{\imath\theta}}dx = \frac{\pi}{n}\left(\sin \frac{m}{n}\pi\right)^{-1} e^{\imath(m/n-1)\theta} \quad 0 < m < n, \ -\pi < \theta < \pi.$$

2. Use Ex. 1 to obtain the following formula due to Euler:
$$\int_0^\infty \frac{x^{m-1}}{x^{2n} + 2x^n\cos\theta + 1}dx = \frac{\pi \sin(1 - m/n)\theta}{n \sin(m\pi/n)\sin\theta}, \quad 0 < m < n, \ -\pi < \theta < \pi.$$

6.7 Two Applications: Error Function and Gauss Sum

In this section we shall give two important applications of residue computations other than merely computing some real integrals. The first one is the so called error integral and the second one is Gauss sum.

The Error Integral: $\displaystyle\int_{-\infty}^{\infty} e^{-(t+z)^2}dt.$

To begin with we know that $e^{x^2} > \frac{x^4}{2!}$ for all $x \in \mathbb{R}$. Therefore, $x^2 e^{-x^2} < \frac{2}{x^2} \to 0$ as $|x| \to \infty$. Therefore, theorem 6.2.1 implies that $\displaystyle\int_{\infty}^{\infty} e^{-t^2}dt$ exists. Fix $w = p + \imath q$ and consider the holomorphic function $f(z) = e^{-z^2}$ and integrate it on the boundary ∂R of the rectangle R with vertices $-r, s, s + \imath q, -r + \imath q$ for r, s.

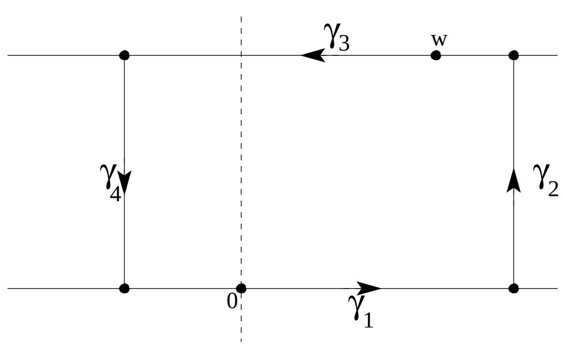


Fig. 34

It is easy to check that
$$\lim_{s\to\infty}\int_{\gamma_2} f(z)dz = 0 = \lim_{r\to\infty}\int_{\gamma_4} f(z)dz.$$

Now we can write $(-\gamma_3)(t) = t + z - r - p \leq t \leq s - p$. Therefore $\int_{-\gamma_3} f(z)dz = \int_{-r-p}^{s-p} e^{(t+z)^2}dt$. By Cauchy's theorem, we have

$$\int_{-\gamma_3} f(z)dz = \int_{-r}^{s} e^{t^2}dt + \int_{\gamma_1+\gamma_2} f(z)dz$$

Upon taking the limit it follows that above Error integral exists and we have

$$\boxed{\int_{\infty}^{\infty} e^{-(t+z)^2}dt = \int_{-\infty}^{\infty} e^{-t^2}dt.} \tag{6.22}$$

We shall now compute the value of $\int_0^\infty e^{-t^2}dt$. Consider the function meromorphic function

$$g(z) = \frac{e^{-z^2}}{1+e^{-2\alpha z}},$$

where $\alpha = \sqrt{\pi}e^{\pi\imath/4}$. Observe that $\alpha^2 = \pi\imath$ and hence α is a period of $e^{-2\alpha z}$. Therefore g has poles precisely at $(-\frac{1}{2}+n)\alpha, n \in \mathbb{Z}$ and all poles are simple. So, we take the rectangle with vertices $-r, s, s+\imath q, -r+\imath q$ where $\alpha = p+\imath q$ which encloses the only pole $z = \alpha/2$.

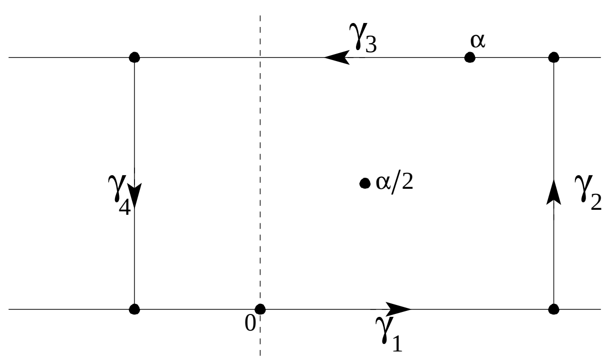


Fig. 35

The residue at this point is $-\frac{\imath}{2\sqrt{\pi}}$. Most important of all is the fact that

$$g(z) - g(z+\alpha) = e^{-z^2}.$$

By Cauchy's theorem, the total integral along the boundary of the rectangle R yields

$$\int_{-r}^{s} e^{-x^2}dx + \int_{\gamma_1+\gamma_3} g(z)dz = \sqrt{\pi}.$$

It is not hard to show that integrals along γ_1 and γ_3 converge to 0 as $r, s \to \infty$. (We leave this to you to verify.) Thus we have

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Because of the evenness of the integrand, this establishes

$$\boxed{\boxed{\int_0^{\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2}.}} \tag{6.23}$$

Gauss Sum: We shall now derive the formula:

$$\boxed{\boxed{\sum_{k=0}^{n-1} e^{\frac{2\pi\imath}{n}k^2} = \frac{1+(-i)^n}{1-i}\sqrt{n}.}} \tag{6.24}$$

popularly known as Gauss sum[1]. Gauss used this to give a proof of quadratic reciprocity. However, we shall not discuss quadratic reciprocity. We begin with an auxiliary lemma:

Lemma 6.7.1 The the function $\phi(a, z) = \frac{e^{az}}{e^z - 1}$ is bounded in $[0, 1] \times U_r$ where

$$U_r = \mathbb{C} \setminus \cup_{n\in\mathbb{Z}} B_r(2\pi\imath n).$$

Proof: Note that the denominator of ϕ is an entire function and has simple zeros at $2\pi\imath n, n \in \mathbb{Z}$. Therefore ϕ is a continuous function on the closure $[0, 1] \times \bar{U}_r$ of $[0, 1] \times U_r$. Also since the denominator is periodic of period $2\pi\imath$ and a is real, it follow that $|\phi|$ is a periodic function of period $2\pi\imath$ in z. Therefore it is enough to show that $|\phi|$ is bounded in $[0, 1] \times S$ where S is the punctured infinite strip

$$S = \{z \in \mathbb{C} \ : \ |\Im z| \leq \pi, |z| \geq r\}.$$

Now clearly $|\phi|$ is bounded on $[0, 1] \times \{z \in S \ : \ |\Re z| \leq 1$. Now for $z = x + \imath y$ with $x > 1$ we have

$$|\phi(a, z)| \leq \frac{e^a x}{|e^z| - 1} = \frac{e^a x}{|e^x| - 1} \leq 2e^{a-1)x} \leq 2.$$

And for $x < -1$, we have

$$|\phi(a, z)| \leq \frac{e^{ax}}{1 - e^{-1}} \leq \frac{1}{1 - e^{-1}}.$$

[1]This formula was entered in his dairy by Gauss in May 1801. The derivation we present is due to L. J. Mordel, 'On a simple summation of the series $\sum_0^{n-1} e^{2s^2\pi\imath/n}$," Messenger of Math. 48(1918), 54-56.

This completes the proof of the lemma. ♠

Now consider the entire function

$$G_n(z) := \sum_{k=0}^{n-1} e^{\frac{2\pi\imath(z+k)^2}{n}}$$

We want to compute $G_n(0)$. Mordel's idea is to consider the meromorphic function

$$M_n(z) := \frac{G_n(z)}{e^{2\pi\imath z}-1},$$

isolate its pole at $z=0$ by the rectangle R_r with vertices $\pm\frac{1}{2}\pm re^{\frac{\imath\pi}{4}}, r>0$ and take the integral on the boundary. The pole $z=0$ is a simple pole and hence the residue of M_n is equal to $G_n(0)/2\pi\imath$. By the residue theorem, we have

$$G_n(0) = \int_{\partial R} M_n(z)dz \tag{6.25}$$

where $\partial R = \gamma_1+\gamma_2+\gamma_3+\gamma_4$ as shown in the figure:

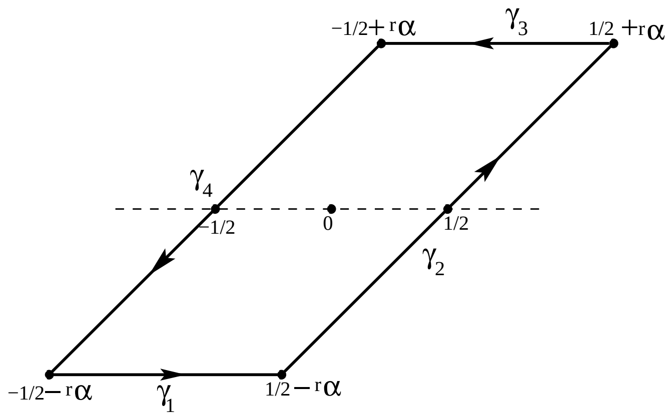


Fig. 36

We shall show that

(a) $\displaystyle\int_{\gamma_1} M_n(z)dz = 0 = \int_{\gamma_3} M_n(z)dz$ and

(b) $\displaystyle\lim_{r\to\infty}\int_{\gamma_2+\gamma_4} M_n(z)dz = (1+(-i)^n)e^{\imath\pi/4}\sqrt{\frac{\pi}{n}}\int_{-\infty}^{\infty} e^{t^2}dt.$

Proof of (a): Put

$$s_\pm(r) = sup\{|M_n(t\pm re^{\frac{\imath\pi}{4}})| \ : \ -1/2\le t\le 1/2\}.$$

Since

$$\int_{\gamma_1} M_n(z)dz = \int_{-1/2}^{1/2} M_n(t-re^{\frac{\imath\pi}{4}})dt; \quad \int_{\gamma_3} M_n(z)dz = \int_{1/2}^{-1/2} M_n(t+re^{\frac{\imath\pi}{4}})dt,$$

by M-L inequality, it is enough to show that $\lim_{r\to\infty} s_\pm(r) = 0$. By lemma 6.7.1, there exists M such that $|(e^{2\pi\imath z}-1)^{-1}| \le M$ for all $z \in \gamma_1 \cup \gamma_3$ and for all r. Since $G_n(z)$ is a finite sum of terms of the form $e^{\frac{2\pi\imath(z+k)^2}{n}}$ we consider

$$\lambda_k(r) = sup\{|e^{\frac{2\pi\imath(z+k)^2}{n}}| \ : \ z = t \pm re^{\frac{\imath\pi}{4}}, -1/2 \le t \le 1/2\}.$$

Then it is enough to prove that $\lim_{r\to\infty} \lambda_k(r) = 0$ for all $0 \le k \le n-1$. For, since $0 \le s_\pm(r) \le M \sum_0^{n-1} \lambda_k(r)$ and hence $\lim_{r\to\infty} s_\pm(r) = 0$. Now

$$|e^{\frac{2\pi\imath(z+k)^2}{n}}| = \Re[2\pi\imath(t \pm re^{\pi\imath/4} + k)^2/n] = \frac{2\pi}{n}(-2r^2 \mp \sqrt{2}(t+k)r).$$

Therefore

$$\lambda_k(r) \le e^{-2\pi r^2} sup\{e^{\mp\sqrt{2}(t+k)2\pi r/n} \ : \ -1/2 \le t \le 1/2\} \le e^{-2\pi r^2 + ckr}$$

for some constant c. The last quantity clearly tends to 0 as $r \to \infty$.

Proof of (b): Since $M_n(z+1) - M_n(z) = e^{\frac{2\pi\imath}{n}z^2}(e^{2\pi\imath z}+1)$ it follows that the integral in (b) is equal to $I_r := \int_{\gamma_4} e^{\frac{2\pi\imath}{n}z^2}(e^{2\pi\imath z}+1)dz$. Now, $\frac{2\pi\imath}{n}z^2 + 2\pi\imath z = \frac{2\pi\imath}{n}(z+\frac{n}{2})^2 - \frac{\pi\imath n}{2}$ and $e^{-\frac{\pi\imath n}{2}} = (-i)^n$, and hence,

$$e^{\frac{2\pi\imath}{n}z^2}(e^{2\pi\imath z}+1) = e^{\frac{2\pi\imath}{n}z^2} + (-i)^n e^{\frac{2\pi}{n}(z+\frac{n}{2})^2}.$$

Putting $z = -1/2 + t\alpha$ where $\alpha = e^{\pi\imath/w}$, and $-r \le t \le r$ and using the fact $\alpha^2 = \imath$, it follows that

$$I_r = \alpha \int_{-r}^{r} e^{\frac{-2\pi\imath}{n}(t-\frac{1}{2\alpha})^2}dt + (-i)^n \int_{-r}^{r} e^{-\frac{2\pi\imath}{n}(t+\frac{1}{2\alpha}(n-1))^2}dt.$$

Upon taking the limit and using the translation invariance of the error integral (6.22), we obtain (b).

6.8 Miscellaneous Exercises to Ch. 6

In the following bunch of exercises, a, b, c are positive real numbers. Evaluate the integrals, using residue method:

1. $\displaystyle\int_0^\infty \frac{dx}{(a+bx^2)^n},\ n \geq 1.$
2. $\displaystyle\int_0^\infty \frac{x^2\,dx}{x^6+1}.$
3. $\displaystyle\int_{-\infty}^\infty \frac{\cos x\,dx}{(x^2+a^2)(x^2+b^2)}.$
4. $\displaystyle\int_{-\infty}^\infty \frac{x\sin x\,dx}{(x+a)^2+b^2},\ a, b > 0.$
5. $\displaystyle\int_0^\infty \frac{\sin x\,dx}{x(x^2-\pi^2)}.$
6. $\displaystyle\int_0^{2\pi} \frac{d\theta}{a+\imath(b\cos\theta + c\sin\theta)}.$
7. $\displaystyle\int_0^{2\pi} \cot\left(\frac{\theta-a-\imath b}{2}\right)\,d\theta.$
8. $\displaystyle\int_0^{2\pi} \frac{\cos 2\theta}{1-2a\cos\theta+a^2},\ a^2 < 1.$
9. $\displaystyle\int_0^{2\pi} \frac{\cos n\theta}{5-3\cos\theta}.$
10. $\displaystyle\int_{-\infty}^\infty \frac{x}{1+x^2}e^{2\imath x}dx.$

11. Show that $\displaystyle\int_0^{2\pi} \frac{d\theta}{1-2a\cos\theta+a^2}\,d\theta = \frac{2\pi}{|a^2-1|},\quad a \neq \pm 1.$

12. With the same hypothesis of theorem 6.3.1, show that the integral $\displaystyle\int_{-\infty}^\infty f(x)e^{\imath ax}\,dx$ exists, as follows: Assume $a > 0$. For $r, s > 0$ let $R(r,s)$ be the square with vertices $-r, s, s+q\imath, -r+q\imath$, where $q = r+s$. Integrate on $\partial R(r,s)$ and show that the integrals on the two vertical sides and the top side tend to 0 as $r, s \longrightarrow \infty$ independently. What do you do for $a < 0$?

13. Using the boundary of the sector $\{z = re^{\imath\theta} : 0 \leq r \leq s,\ 0 \leq \theta \leq \phi\}$ where $\phi = 2\pi/n$, show that for integers $0 < m < n$, we have,

$$\int_0^\infty \frac{x^{m-1}}{1+x^n}dx = \frac{\pi}{n\sin\frac{m\pi}{n}}.$$

14. Let f be a meromorphic function and γ be a contour passing through some of the singularities of f. We wish to extend the definition of the integral to cover some of these situations. Let $\xi_j, j = 1, 2, \ldots, n$ be such singularities of the contour. Choose $\epsilon > 0$ sufficiently small so that $B_\epsilon(\xi_j)$ are all disjoint. We can also assume that inside each $B_\epsilon(\xi_j)$, the portion of the contour consists of two smooth arcs, joined at ξ_j. Now consider the portion of the contour that lies outside $\cup_j B_\epsilon(\xi_j)$, denote it by Γ_ϵ. Clearly $\displaystyle\int_{\Gamma_\epsilon} f(z)dz$ exists. Define

$$PV\left(\int_\gamma f(z)dz\right) = \lim_{\epsilon\longrightarrow 0}\int_{\gamma_\epsilon} f(z)dz \tag{6.26}$$

if it exists.

(a) Let A be the sector defined by

$$A = \{z = re^{\imath\theta} \ : \ 0 \le r \le s, \ \theta_1 \le \theta \le \theta_2\}.$$

Take τ to be the boundary of A traced in counter clockwise direction. Show that $PV\left(\int_\tau \frac{dz}{z}\right) = \imath(\theta_2 - \theta_1)$.

(b) Deduce that if f has a simple pole at 0 and s is sufficiently small, then with τ as in a), we have,

$$PV\left(\int_\tau f(z)dz\right) = \imath(\theta_2 - \theta_1)R_0(f).$$

(c) Extend the result (b) to the case when τ is replaced by a curve which is similar to τ except that instead of two line segments at 0 it has a two arcs which are tangents to these two line segments.

(d) Prove that if all ξ_j are simple poles then 6.26 exists. Indeed if θ_j is the angle subtended by the *left-side tangent* with the *right side tangent* to γ at ξ_j measured in the appropriate direction, then

$$PV\left(\int_\gamma f(z)dz\right) = 2\pi\imath\left(\sum_{a\in A}\eta(\gamma, a)R_a(f)\right) + \sum_{j=1}^{n}\theta_j R_{\xi_j}(f),$$

where A is the set of all poles not lying on γ.

15. Let $z_j, \ j = 1, 2, 3, 4$ form the vertices of a rectangle R. For any complex numbers $a_j, j = 1, 2, 3, 4$ let

$$f(z) = \sum_{j=1}^{4} \frac{a_j}{z - z_j}.$$

Find $PV\left(\int_{\partial R} f(z)\, dz\right)$ where ∂R is the boundary of the rectangle traced in the counter clockwise sense.

16. Let $z_1 = 1 + \imath, z_2 = 1 - \imath, z_3 = -1 - \imath, z_4 = 1 - \imath$ in the above exercise. Find $\int_{\partial R} \frac{1}{z^4 + 4}\, dz$.

Chapter 7

Local And Global Properties

7.1 Schwarz's Lemma

A combination of maximum modulus principle with Riemann's removable singularity yields the following geometrically important result.

Theorem 7.1.1 Schwarz's Lemma *Let $f : \mathbb{D} \to \bar{\mathbb{D}}$ be a holomorphic function such that $f(0) = 0$. Then $|f(z)| \leq |z|$ and $|f'(0)| \leq 1$. Further, the following three conditions are equivalent:*

(i) *there exists $z_0 \neq 0$ with $|z_0| < 1$ and $|f(z_0)| = |z_0|$.*

(ii) $|f'(0)| = 1$.

(iii) $f(z) = cz$ *for some* $|c| = 1$.

Proof: Take

$$g(z) = \begin{cases} f(z)/z, & z \neq 0, \\ f'(0), & z = 0. \end{cases}$$

Clearly g is holomorphic in $\mathbb{D} \setminus \{0\}$. Moreover, by the very definition of $f'(0)$, we have $\lim_{z\to 0} g(z) = f'(0) = g(0)$. This implies that 0 is a removable singularity of g and hence then $g(z)$ is holomorphic in $\mathbb{D}$. Also for $0 < r < 1$ and $|z| = r$, we have, $|g(z)| = |f(z)|/|z| \leq 1/r$. By the maximum principle, $|g(z)| < 1/r$ for $|z| < r$. Letting $r \to 1$, we obtain $|g(z)| \leq 1$. Hence $|f(z)| \leq |z|$. Also, since $f'(0) = g(0)$, we have, $|f'(0)| \leq 1$.

To prove the latter part, suppose (i) holds. Then it follows that $|g(z_0)| = 1$ for an interior point z_0 of the domain, and hence by maximum principle, $g(z)$ must be a constant of modulus 1. This implies (iii). Likewise (ii) implies that $|g(0)| = |f'(0)| = 1$

and hence by maximum principle etc., we get (iii). The implications (iii) $\Longrightarrow$ (i) & (iii) $\Longrightarrow$ (ii) are obvious. This completes the proof. ♠

Remark 7.1.1 The conditions stated in the theorem 7.1.1 are not all mandatory. Neither the conclusion is the strongest. We can improve upon both of them in several ways by using this theorem itself. Roughly speaking the theme underlying Schwarz's lemma is that starting with a holomorphic function which has a rough fixed bound on a bounded set, it is possible to get a stronger 'variable bound' for f. Here are two such instances. For more, see exercise 2 below.

Example 7.1.1 Suppose f is holomorphic in the disc $|z| < R$ and maps it inside the disc $|w| < M$, and $f(0) = 0$. Then, we can first take the map $T(z) = Rz$, follow it by f and then take $S(z) = z/M$ to get a mapping $g = S \circ f \circ T : \mathbb{D} \longrightarrow \mathbb{D}$ such that $g(0) = 0$. Applying Schwarz's lemma, we get, $|g(w)| \leq |w|$ for all $w \in \mathbb{D}$. Replacing w by Rz, this is the same as

$$|f(z)| \leq \frac{M}{R}|z|,$$

which is an improvement on the data $|f(z)| < M$. Thus, the hypotheses in Schwarz's lemma are not all necessary in a sense.

Example 7.1.2 Given a holomorphic map $f : \mathbb{D} \to \mathbb{D}$, what can we say when $f(0) \neq 0$? Remember the mapping (see theorem 3.7.5)

$$L_a : z \mapsto \frac{z-a}{1-\bar{a}z}$$

for $a \in \mathbb{D}$? This comes to our rescue now. It maps the unit disc onto itself and maps a to 0. All that we have to do is to compose f with $L_{f(0)}$. The map $g = L_{f(0)} \circ f$ fits the bill of Schwarz's lemma. So we may conclude that

$$|f(z) - f(0)| \leq (|1 - \overline{f(0)}f(z)|)|z|, \quad \forall\ z \in \mathbb{D}. \tag{7.1}$$

Dividing out by z and taking the limit as $z \to 0$, we get

$$|f'(0)| \leq |1 - |f(0)|^2| = 1 - |f(0)|^2. \tag{7.2}$$

Both (7.1),(7.2) are stronger than the corresponding conclusions in theorem 7.1.1.

As an immediate corollary, we obtain a complete description of the automorphisms of the unit disc.

Theorem 7.1.2 *The set of all automorphisms of the unit disc $\mathbb{D}$ is in one-to-one correspondence with the set of fractional linear transformations of the form*

$$\boxed{f(z) = c\frac{z+b}{1+\bar{b}z}, \quad b \in \mathbb{D}, \ c \in \mathbb{S}^1.} \tag{7.3}$$

Indeed, $c = f'(0)/(1-|f(0)|^2) = f'(0)/|f'(0)|$, and $b = \bar{c}f(0) = -f^{-1}(0)$.

Proof: In view of theorem 3.7.5, we need to prove that every automorphism f of $\mathbb{D}$ is a fractional linear transformation.

First assume that $f(0) = 0$. Put $g = f^{-1}$. Then $g(0) = 0$ and by Schwarz's lemma, we have $|f'(0)| \leq 1$, and $|g'(0)| \leq 1$. But then by chain rule, $|g'(0)| = |1/f'(0)| \geq 1$. Hence, $|f'(0)| = 1$. (Alternatively, we have $|z| = |g \circ f(z)| \leq |f(z)| \leq |z|$ for all z.) Again by (later part of) Schwarz's lemma, $f(z) = cz$ for some $c = e^{\imath\theta}$. That is, f is a rotation through an angle θ. In the general case, consider

$$h(z) := L_{f(0)}(z) := \frac{f(z) - f(0)}{1 - \overline{f(0)}f(z)}$$

as in the above example. Then $h(0) = 0$ and hence by the previous case, we get $h(z) = cz$ where $c = h'(0)$ is of modulus 1. Clearly

$$c = h'(0) = \frac{f'(0)}{1 - |f(0)|^2}.$$

Thus

$$f(z) = \frac{cz + f(0)}{1 + \overline{f(0)}cz} = c\frac{z + \bar{c}f(0)}{1 + \overline{\bar{c}f(0)}z} = c\frac{z+b}{1+\bar{b}z}.$$

where $c = f'(0)/(1 - |f(0)|^2)$ and $b = \bar{c}f(0)$. Since c is of modulus 1 as well, it follows that $c = \frac{f'(0)}{|f'(0)|}$. ♠

Remark 7.1.2 For a geometric interpretation of Schwarz lemma see the exercise 3 below. For some beautiful applications and more see [[**?**]].

Exercise 7.1

1. Show that the automorphisms of the unit disc are in one-to-one correspondence with the set of fractional linear transformations of the form $\dfrac{az+b}{\bar{b}z+\bar{a}}$, where $|a|^2 - |b|^2 = 1$.

2. Suppose $f(z)$ is holomorphic in the disc $|z| < R$ and maps it inside the disc $|f(z)| < M$. Prove that

$$\left|\frac{M(f(z_1)-f(z_2))}{M^2-\overline{f(z_2)}f(z_1)}\right| \leq \left|\frac{R(z_1-z_2)}{R^2-\overline{z_2}z_1}\right|, \; z_1, z_2 \in B_R(0).$$

[Hint: Use exercise 3.7.2] Deduce that

$$\frac{M|f'(z)|}{M^2-|f(z)|^2} \leq \frac{R}{R^2-|z|^2}, \text{ for all } |z| < R.$$

The case $R = M = 1$ gives the so called Schwarz-Pick theorem. viz.,

$$\left|\frac{(f(z_1)-f(z_2))}{1-\overline{f(z_2)}f(z_1)}\right| \leq \left|\frac{(z_1-z_2)}{1-\overline{z_2}z_1}\right|, \; z_1, z_2 \in \mathbb{D}. \tag{7.4}$$

3. Define $d : \mathbb{D} \times \mathbb{D} \to [0, 1)$ by the formula

$$d(z_1, z_2) = \left|\frac{z_1-z_2}{1-z_1\bar{z}_2}\right|. \tag{7.5}$$

Show that
(i) d is invariant under rotation and Möbius transformation $\mathbb{D}$ i.e., if $h : \mathbb{D} \to \mathbb{D}$ is either a rotation or a Möbius transformation then $d(h(z_1), h(z_2)) = d(z_1, z_2)$.
(ii) d is a metric on $\mathbb{D}$, i.e., d is symmetric, $d(z_1, z_2) = 0$ iff $z_1 = z_2$ and d satisfies the triangle inequality:

$$d(z_1, z_3) \leq d(z_1, z_2) + d(z_2, z_3), \quad z_1, z_2, z_3 \in \mathbb{D}.$$

Now Schwarz-Pick Theorem (7.4) has the following geometric interpretation: *every holomorphic self-mapping of the unit disc decreases the noneuclidean distance d.*

4. Suppose $f : \boldsymbol{H} \longrightarrow \boldsymbol{H}$ is a holomorphic mapping. Then show that

$$\left|\frac{f(z)-f(w)}{f(z)-\overline{f(w)}}\right| \leq \left|\frac{z-w}{z-\overline{w}}\right| \;\; \forall\, z, w \in \boldsymbol{H}$$

and hence deduce that

$$|f'(z)| \leq \frac{\Im(f(z))}{\Im z}.$$

5. Show that if $|f(z)| < 1$ for $|z| < 1$ and $f(z)$ has a zero of order n at 0 then
   (i) $|f(z)| < |z|^n$ for all $|z| < 1$.
   (ii) $|f^{(n)}(0)| \leq n!$.
   Also, show that equality occurs in any one of the above iff $f(z) = cz^n$ with $|c| = 1$.

6. Let f be an entire function. Suppose there is a constant a and an integer n such that $|f(z)| \leq a|z|^n, \forall z \in \mathbb{C}$. Then show that $f(z) = cz^n$, for some $c \in \mathbb{S}^1$.

7. Show that every automorphisms of the upper-half plane is a flt of the form
$$z \mapsto \frac{az+b}{cz+d}, \quad ad - bc = 1, \quad a, b, c, d \in \mathbb{R}.$$
   [Hint: see Ex. 3.7.6.]

8. Suppose $f : \mathbb{D} \to \mathbb{D}$ is a holomorphic mapping having a continuous extension on the boundary such that $f(\partial\mathbb{D}) \subset \partial\mathbb{D}$. Show that f is a rational function.

7.2 Local Mapping

We begin with a deep application of argument principle.

Theorem 7.2.1 Existence of Solutions *Let f be holomorphic in a neighborhood of z_0, $f(z_0) = w_0$ and let $f(z) - w_0$ have a zero of order n at z_0. Then there exist $\epsilon > 0$ and $\delta > 0$ such that if $0 < |a - w_0| < \delta$, then*

$$f(z) = a$$

has exactly n simple (i.e., distinct) solutions in the disc $|z - z_0| < \epsilon$.

Proof: Choose $\epsilon > 0$ such that f is defined and holomorphic in $|z - z_0| < 2\epsilon$ and the equation

$$f(z) = w_0$$

has no solution in $0 < |z - z_0| < 2\epsilon$. This is possible because zeros of a holomorphic function are isolated. For the same reason, we can further demand that

$$f'(z) \neq 0 \text{ in } 0 < |z - z_0| < 2\epsilon$$

also. Let γ be the circle

$$\gamma(\theta) = z_0 + \epsilon e^{\imath\theta}, \; 0 \leq \theta \leq 2\pi$$

and let $\Gamma = f \circ \gamma$. Observe that $w_0 \notin \Gamma$. So choose $\delta > 0$ such that $B_\delta(w_0) \cap \Gamma = \emptyset$.

Now, the mapping $a \mapsto \eta(\Gamma, a)$ is defined and continuous on the entire disc $B_\delta(w_0)$. Being an integer valued function, it is a constant. Therefore, $\eta(\Gamma, a) = \eta(\Gamma, w_0)$ for all $a \in B_\delta(w_0)$. On the other hand, we have $\eta(\Gamma, w_0) = \eta(\gamma, z_0)n = n$, since $z = z_0$ is a root of $f(z) = w_0$ of order n and there are no other roots inside γ. Hence $\eta(\Gamma, a) = n$ for all $a \in B_\delta(w_0)$.

Now, let $z_1, \ldots, z_m$ be the roots of $f(z) = a$ which lie inside γ. Since $f'(z_i) \neq 0$ each z_i is a simple root. Hence from remark 5.7.2, it follows that $m = n$. ♠

We can now deduce several important results with little effort. First of all:

Corollary 7.2.1 Let $f : \Omega \to \mathbb{C}$ be a holomorphic function, $z_0 \in \Omega$ be such that $f'(z_0) = 0$. Then for every $\delta > 0$ such that $B_\delta(z_0) \subset \Omega$, f is not injective on $B_\delta(z_0) \setminus \{z_0\}$.

Next, we shall give another proof of the open mapping theorem.

Theorem 7.2.2 Open Mapping Theorem *A non-constant holomorphic function on an open set is an* open mapping.

Proof: Let $f : \Omega \longrightarrow \mathbb{C}$ be a holomorphic function and let U be any open subset of Ω. We must show that $f(U)$ is an open set in $\mathbb{C}$. Let $w_0 \in f(U)$ be any arbitrary point, say, $w_0 = f(z_0)$, $z_0 \in U$. Choose $\epsilon > 0$ and $\delta > 0$ as in the above theorem. Since z_0 is a root of $f(z) - w_0$ of some order $l > 0$, it follows that for all $w \in B_\delta(w_0)$, we can find l points in $B_\epsilon(z_0)$ which are mapped to w by f. In particular, it follows that $B_\epsilon(w_0) \subset f(U)$. Since this happens for each point $w_0 \in f(U)$, it follows that $f(U)$ is open. ♠

Theorem 7.2.3 Inverse Function Theorem *Let f be holomorphic on an open set Ω. If $f'(z_0) \neq 0$, for some $z_0 \in \Omega$, then there exists a neighborhood of z_0, on which f has a holomorphic inverse.*

Proof: Since $f'(z_0) \neq 0$, putting $w_0 = f(z_0)$, it follows that z_0 is a simple root of $f(z) - w_0 = 0$ in a neighborhood of z_0. So there exists a neighborhood Ω_1 of z_0 mapped onto a neighborhood Ω_2 of w_0 by f and the map $f : \Omega_1 \longrightarrow \Omega_2$ is injective. Now apply the branch theorem 5.2.2. ♠

Combining the open mapping theorem with theorem 1.5.7, we obtain another proof of :

Theorem 7.2.4 Maximum Modulus Principle: *If $f(z)$ is a non-constant holomorphic function in an open set Ω, then its absolute value $|f(z)|$ has no maximum in Ω.*

Argument Principle can be applied to prove existence theorems of various kind. Here are two applications.

Theorem 7.2.5 Rouchè's[1] Theorem : *Let f, g be holomorphic in an open set containing the closure $\bar{D}$ of a disc D and satisfy the inequality*

$$|f(z) - g(z)| < |g(z)| \ \ \forall\ z \in \partial D.$$

Then f and g have same number of zeroes inside D.

Proof: Clearly $f(z) \neq 0$ and $g(z) \neq 0$ on ∂D. Therefore the inequality yields

$$\left| \frac{f(z)}{g(z)} - 1 \right| < 1$$

for all points on ∂D. Hence the function $F = f/g$ maps the ∂D onto a closed contour Γ inside the ball $B_1(1)$ and so Γ does not wind around 0. Therefore by the Argument Principle we have, $0 = \eta(\Gamma, 0) =$ the number of zeros of F minus the number of poles of F enclosed by ∂D $=$ the number of zeros of f minus the number of zeros of g inside D. ♠

Example 7.2.1 *As an interesting application let us prove the following:*
Let f be holomorphic on $\overline{\mathbb{D}}$ and $|f(z)| < 1$ for $|z| = 1$. Then f has exactly one fixed point inside $\mathbb{D}$.

We apply Rouche's principle to the functions $f(z)-z$ and $-z$. Then $|f(z)-z-(-z)| = |f(z)| < |-z|$, for all $|z| = 1$. Hence $f(z) - z$ and $-z$ have the same number of zeros inside $\mathbb{D}$. Since $-z$ has precisely one zero, we are done.

We shall now derive an important corollary of the argument principle about the elementary symmetric functions on the solution sets. See exercise 2.2.2.

Theorem 7.2.6 Holomorphicity of Elementary Symmetric Functions: *Let σ_i, $i = 1, \ldots, n$ be the elementary symmetric functions in n variables. Let f be holomorphic at $z_0, f(z_0) = w_0$ and let $f(z) - w_0$ have a zero of order n. Then there exists $\epsilon > 0$ and $\delta > 0$*

[1] Eugène Rouché (1832-1910) was a French mathematician. He proved this theorem in 1862.

such that for each w in $B_\delta(w_0)$, $f(z)-w$ has precisely n solutions $z_j(w), j = 1, 2, \ldots, n$, in $B_\epsilon(z_0)$. Moreover, any symmetric function on $\{z_1(w), z_2(w), \ldots, z_n(w)\}$ is well-defined single valued and holomorphic function of w inside $B_\delta(w_0)$.

Proof: In view of theorem 7.2.1, the first part of the theorem has been already seen. The only thing that we need to prove now is the single valuedness and holomorphicity of the symmetric functions. Observe that each of the $z_j(w)$ may not be even continuous, since we may have messed up in labeling the distinct roots as w varies. Since the value of any symmetric function of these n roots $z_j(w)$ does not depend on the labeling, the well definedness follows.

Recall that any symmetric function is a polynomial function of the elementary symmetric functions. Hence it is enough to show that the elementary symmetric functions

$$s_i(w) = \sigma_i(z_1(w), \ldots, z_n(w))$$

are holomorphic. Now recall that if $\rho_k(z) = \sum_j z_j^k$ are the power-sum functions then we have the Newton's identities:

$$\rho_k - \sigma_1\rho_{k-1} + \cdots + (-1)^{k-1}\sigma_{k-1}\rho_1 + (-1)^k k\sigma_k = 0.$$

In particular, it follows that each σ_k is a polynomial function of the power-sum functions ρ_j's. Hence, it suffices to prove that each power-sum function

$$r_k(w) = \rho_k(z_1(w), \ldots, z_n(w)) = \sum_j z_j^k(w)$$

is holomorphic. All that we do now is to fix $w \in B_\delta(w_0)$, apply Logarithmic Residue Theorem 5.7.3 to the function $f(z) - w$ with $g(z) = z^k$ (for each fixed k) and $\gamma := \{|z-z_0| = \epsilon\}$. The zeros of $f(z)-w$ inside $|z-z_0| = \epsilon$ are precisely $z_j(w), j = 1, 2, \ldots, n$. And $f(z) - w$ does not have any poles. Since $\eta(\gamma, z_j(w)) = 1$ for each j, we to obtain,

$$\boxed{\frac{1}{2\pi\imath}\int_{|z-z_0|=\epsilon} \frac{f'(z)}{f(z)-w} z^k\,dz = \sum_j g(z_j(w)) = \sum_j z_j^k(w) = r_k(w)} \tag{7.6}$$

Thus, we have indeed displayed the power-sums of the roots by integral formulae and hence the holomorphicity of these functions follow. ♠

Finally as a special case of (7.6), we shall obtain an integral formula for the inverse function of a holomorphic function when it exists. So, assume that f is holomorphic and injective in an open set Ω. Let $z_0 \in \Omega$ and $\epsilon > 0$ be such that $\overline{B_\epsilon(z_0)} \subset \Omega$. Then in the above theorem, we can take $n = 1$ because $f(z) - w_0$ has a root of multiplicity 1. Also, the elementary symmetric function $r_1(w) = z_1(w) = f^{-1}(w)$. Hence the formula (7.6) yields the following:

Theorem 7.2.7 Integral formula for the inverse function : *Let f be an holomorphic function at z_0, $f(z_0) = w_0$ and suppose that $f(z) - w_0$ has a simple zero at $z = z_0$. Then for all sufficiently small $\epsilon > 0$ there exists a $\delta > 0$ such that for $|w - w_0| < \delta$ we have*

$$f^{-1}(w) = \frac{1}{2\pi\imath} \int_{|z-z_0|=\epsilon} \frac{f'(z)z}{f(z) - w} dz. \tag{7.7}$$

Exercise 7.2

1. Let Ω be a bounded open set, f and g be non vanishing continuous functions on $\overline{\Omega}$ and holomorphic in Ω. Suppose that $|f(z)| = |g(z)|$ for all $z \in \partial\Omega$. Then show that $f(z) = e^{\imath\theta} g(z)$ for all $z \in \overline{\Omega}$ and for some fixed θ.

2. Let f be a non constant holomorphic function in a disc $B_R(0)$ and for $0 \le r < R$, let $M(r) = sup\{|f(z)| : |z| = r\}$. Show that $M(r)$ is strictly monotonically increasing continuous function of r.

3. Determine the number of zeros of the following polynomials inside the unit circle.
   (a) $T^7 - 4T^3 + T - 1$.
   (b) $T^6 - 5T^4 + T^3 - 2T$.
   (c) $2T^4 - 2T^3 + 2T^2 + 2T - 9$.

4. Show that the equation $6T^5 + 2T^2 + T - 1 = 0$ has all its solutions inside $|z| \le 1$. Can you say the same about the solutions of $6T^5 - 2T^2 - T - 1 = 0$? Try to generalize this result about solution set of polynomials. Compare your answer with ex.6 of 7.5.

7.3 Homotopy and Simple Connectivity

On a simple open arc, there is 'essentially' only one way to go from one point to another. In contrast, on a circle, there are at least two different ways to do this. As we have already seen, one can interpret the word 'to go' here to mean 'to communicate' or 'to connect by a path'. Thus the first case could be referred to as 'simple connectivity' and the later as 'multi-connectivity'. This is how the originators of this notion must have thought as the words used by them indicate. In modern times, these notions are made to work in a larger context and hence a certain abstract, more rigorous and (hence) dry definitions have been adopted in the study of Algebraic Topology using the machinery of the fundamental group. We shall not take full recourse to that here, whereas we shall introduce the concept of homotopy and 'a correct' modern definition of simple connectedness. Classically the approach for simple connectivity came through the properties of integrals on them. We prefer to call this 'homological' simple connectivity.

It is time now that you strengthen your topological background. For instance, have you solved all the exercises in 1.6? We will need some of these results such as exercise 4 from 1.6.

Definition 7.3.1 Let $\gamma_j : [0,1] \to \gamma,\ j = 0, 1$, be any two paths with the same initial and terminal points:

$$\gamma_0(0) = \gamma_1(0) = c_0; \gamma_0(1) = \gamma_1(1) = c_1.$$

We say γ_0, γ_1 are *path-homotopic* to each other in Ω and express this by writing $\gamma_0 \sim \gamma_1$ if there exists a continuous map $H : I \times I \to \Omega$ such that

$$H(t,j) = \gamma_j(t);\ H(j,s) = \gamma_0(j) = \gamma_1(j) = c_j, \quad j = 0,1;\ 0 \le t \le 1;\ 0 \le s \le 1.$$

H is called a path-homotopy from γ_0 to γ_1.

If $c_0 = c_1 = a$ that is when both the paths are loops passing through a, the above path-homotopy gives a *'loop homotopy'* of loops based at a. If γ_0 happens to be the constant loop, we say the loop γ_1 is *'null-homotopic.'*

The importance of this notion for us lies in the local-to-global result that we can now prove. We need just one more definition:

Definition 7.3.2 A differential 1-form $pdx + qdy$ on a domain Ω is said to be locally exact if for each point $z \in \Omega$, there is an open disc around z contained in Ω on which $pdx + qdy$ is exact.

Remark 7.3.1 Recall that $pdx + qdy$ is exact means that there is a smooth function $f(x, y)$ such that $f_x = p$ and $f_y = q$. Further, if p and q are continuously differentiable, then we see that in order that $pdx + qdy$ is locally exact, it is necessary that $p_y = q_x$, i.e., the form $pdx + qdy$ is a *closed form.* Combining corollary 4.2.1 with Green's theorem applied to triangles, it follows that any closed 1-form is locally exact. (However, in the sequel, we do not need this result.) We shall prove:

Proposition 7.3.1 Let $pdx+qdy$ be a locally exact 1-form in a domain Ω. Let $\gamma_j, \ \ j = 1, 2$ be any two contours in Ω which are path homotopic in Ω. Then

$$\int_{\gamma_0} pdx + qdy = \int_{\gamma_1} pdx + qdy.$$

By Cauchy's theorem for discs, we know that for a holomorphic function f, fdz is locally exact. Therefore proposition 7.3.1 yields:

Theorem 7.3.1 Homotopy Invariance of Integrals *Let f be a holomorphic mapping on a domain Ω, and γ_j be any two contours in γ which are path homotopic in Ω. Then*

$$\int_{\gamma_0} fdz = \int_{\gamma_1} fdz.$$

Corollary 7.3.1 Let γ be a null-homotopic contour in Ω. Then for every holomorphic function f on Ω, we have

$$\int_{\gamma} fdz = 0.$$

Toward the proof of proposition 7.3.1, for the first time we shall now need the compactness notion in a sense in which it is meant viz., every open cover of a compact space admits a finite subcover. In the case of metric spaces, the accompanying result that we need is Lebesgue Covering Lemma:

Theorem 7.3.2 Lebesgue Covering Lemma *To each open cover of a compact metric space, there exists $r > 0$ such that any ball of radius r is contained in some member of the cover.*

We shall assume this result. The reader may refer to chapter 6 in Joshi's book [J].

Proof of the proposition 7.3.1: The idea is that the homotopy H defines a 'continuous family' $\{\gamma_s \ : \ 0 \leq s \leq 1\}$ of paths beginning with γ_0 and ending with γ_1 and each of them having the same end points. The claim is that for all these paths the integral $\int_{\gamma_s} pdx + qdy$ takes the same value. Unfortunately, even to make sense out of this claim there is a technical snag: the intermediary paths $\gamma_s, 0 < s < 1$, may not be piecewise smooth. So, we use these paths as a device to obtain finitely many contours (made up of line segments) with the same end-points so that from one contour to the other the value of the integral does not change. Let us now get into the technical details.

Choose a path homotopy $H : I \times I \to \Omega$ from γ_0 to γ_1. Cover $H(I \times I)$ by finitely many open balls $\{B_\alpha\}$ which are completely contained in Ω and on each of which the differential is exact. Choose $r > 0$ so that every ball of radius r in $I \times I$ is contained in some member of $\{H^{-1}(B_\alpha)\}$ (such a $r > 0$ exists by Lebesgue Covering Lemma). Choose an integer n such that $1/n < r/2$. Cut $I \times I$ into n^2 squares of size $\frac{1}{n} \times \frac{1}{n}$. Since each little square is of diameter $< r$, it follows that each of these little squares is contained in one of the members of $\{H^{-1}(B_\alpha)\}$.

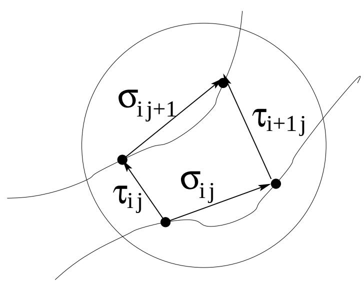


Fig. 37

Let $a_{i,j} = H(\frac{i}{n}, \frac{j}{n})$. Let τ_{ij} be the line segment joining $a_{i,j}$ and $a_{i,j+1}$ for $0 < i < n, 0 \leq j < n$. Observe that $a_{0,j} = \gamma_0(0) = c_0, a_{n,j} = \gamma_0(1) = c_1$ for all j. Let σ_{ij} be the line segment joining $a_{i,j}$ to $a_{i+1,j}$ for $0 \leq i < n, 0 < j < n$. Let $\sigma_{i,0}, \sigma_{i,n}$ denote the contour γ_0 (respectively γ_1 restricted to the interval $[\frac{i}{n}, \frac{i+1}{n}], 0 \leq i < n$.

Let P_{ij} denote the closed contour

$$P_{ij} := \sigma_{ij} \star \tau_{i+1,j} \star \sigma_{i,j+1}^{-1} \star \tau_{i,j}^{-1}.$$

Clearly, P_{ij} is a closed contour and is contained in one of the discs B_α contained in Ω. Therefore

$$\int_{P_{ij}} pdx + qdy = 0; \quad i, j = 0, 1, \ldots, n-1. \tag{7.8}$$

Therefore

$$\sum_{ij} \int_{P_{ij}} pdx + qdy = 0. \tag{7.9}$$

As usual, all the integrals along intermediate line segments cancel in pairs and we are left with the integrals on $H(\partial(I \times I))$. This is nothing but $\gamma_0 \star c_1 \star \gamma_1^{-1} \star c_0$ where c_0, c_1

are constant paths. Therefore

$$\int_{\gamma_0} pdx + qdy = \int_{\gamma_1} pdx + qdy.$$

This completes the proof of the proposition 7.3.1 as well as those of theorem 7.3.1 and Corollary 7.3.1. ♠

Definition 7.3.3 Let $\Omega \subset \mathbb{C}$ be a domain. We say Ω is simply connected, if every closed path in Ω is null-homotopic in Ω.

Remark 7.3.2

1. Let $f : U \to V$ is a continuous function. If $\gamma : I \to U$ is a loop which is null homotopic in U, then it follows that $f \circ \gamma$ is null homotopic in V. Indeed, if $H : I \times I \to U$ is a null-homotopy of γ, then $f \circ H$ gives a null-homotopy of $f \circ \gamma$.

2. As a consequence, it follows that simple connectivity is a topological invariant i.e., if $f : U \to V$ is a homeomorphism then U is simply connected iff V is simply connected.

3. It is not very difficult to see that if Ω is a domain and $z_0 \in \Omega$ is such that every loop in Ω based at z_0 is null homotopic then Ω is simply connected. However, we need not use this.

As an immediate corollary, in view of theorem 4.2.1, we obtain:

Theorem 7.3.3 *A locally exact (equivalently, a closed) differential 1-form on a simply connected domain is exact.*

Remark 7.3.3 The entire plane is simply connected. Indeed any convex domain in $\mathbb{C}$ is simply connected. Since simple connectivity is a topological invariant property, any domain which is homeomorphic to a convex domain is simply connected. At this stage, we do not know any other way to see more examples of simply connected domains. Neither we have any tools to test whether a given domain is simply connected or not. The above corollary fills this gap to a certain extent. Let us restate it as:

Theorem 7.3.4 Cauchy's Theorem: Homotopy Version *Let Ω be a simply connected domain. Then for every closed contour γ in Ω and every holomorphic function f, $\displaystyle\int_\gamma f dz = 0$.*

Remark 7.3.4 Thus, if we find one holomorphic function f on Ω and one closed contour in Ω such that $\int_\gamma f(z)dz \neq 0$, then we know that Ω is not simply connected. Thus $\mathbb{C}\setminus\{0\}$ is not simply connected because $1/z$ is holomorphic on it and its integral on the unit circle is $2\pi\imath$. Indeed, with this method, we are sure that given any domain Ω and a non empty, finite subset A of Ω, the domain $\Omega \setminus A$ is not simply connected.

We still do not know any sure method to determine whether a given domain is simply connected or not. The following theorem takes us quite close to settling this problem.

Theorem 7.3.5 *Let Ω be a domain in $\mathbb{C}$. Consider the following statements:*
(0) *Either $\Omega = \mathbb{C}$, or Ω is biholomorphic to the opne unit disc $\mathbb{D}$.*
(i) *Ω is simply connected.*
(ii) *For every closed contour γ in Ω and every holomorphic function f, $\int_\gamma f dz = 0$.*
(iii) *Every holomorphic function in Ω has a primitive.*
(iv) *Every holomorphic function on Ω which never vanishes on Ω has a holomorphic logarithm, i.e., there exists a holomorphic function $g : \Omega \to \mathbb{C}$ such that $\exp(g(z)) = f(z), z \in \Omega$.*
(v) *For every point $a \in \mathbb{C} \setminus \Omega$, and every closed contour γ in Ω, we have, $\eta(\gamma, a) = 0$.*
(vi) For evey point $a \in \mathbb{C} \setminus \Omega$, $\sqrt{z-a}$ has a (continuous) branch all over Ω.
We have (0) $\Longrightarrow$ (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$(v) *and (iv) $\Longrightarrow$ (vi).*

Proof: (0) $\Longrightarrow$ (i) See remark 7.3.2.2.
(i) $\Longrightarrow$ (ii): This is the statement of theorem 7.3.4.
(ii) $\Longrightarrow$ (iii) : This is the statement of the primitive existence theorem (see cor 4.2.2) for each fixed holomorphic function f.
(iii) $\Longrightarrow$ (iv): Apply (iii) to $h = f'/f$ to obtain g such that $g' = f'/f$. Then

$$(\exp(-g)f)' = -\exp(-g)g'f + \exp(-g)f' = \exp(-g)(-f' + f') = 0.$$

Therefore $k(\exp(g)) = f$ for some constant $k \neq 0$. By choosing g such that $\exp(g)$ and f coincide at a point, we get $\exp(g) = f$.
(iv) $\Longrightarrow$ (v): The function $f(z) = z - a$ does not vanish on Ω. Therefore we have a holomorphic g on Ω such that $exp(g) = f$. But then

$$\eta(\gamma, a) = \frac{1}{2\pi\imath}\int_\gamma \frac{dz}{z-a} = \frac{1}{2\pi\imath}\int_\gamma dg = \frac{1}{2\pi\imath}(g(\gamma(b)) - g(\gamma(a))) = 0.$$

(iv) $\Longrightarrow$ (vi): $f(z) = z - a$ does not vanish on Ω. Therefore we have a holomorphic g on Ω such that $\exp(g) = f$. Take $h(z) = \exp(g(z)/2)$ to see that $h^2(z) = f(z)$.

This completes the proof of the theorem. ♠

Remark 7.3.5 Indeed, it is true that all these statements are equivalent. Obviously the difficulty is in proving (vi) $\Longrightarrow$ (0) or (v) $\Longrightarrow$ (i) or for that matter in proving (ii) $\Longrightarrow$ (i).

We shall actually prove (vi) $\Longrightarrow$ (0), in the next chapter and this goes under the name Riemann mapping theorem. In particuar, this will complete the proof of equivalence of all the statements in the above theorem except (v).

So, we shall now launch a programme which will enable us to prove (v) $\Longrightarrow$ (ii). Classically, and in many books even today, simple connectivity is defined by condition (v). Then the statement '(v) implies (ii)' is known as Homology form of Cauchy's theorem. The property (v) can be termed as 'homological simple connectivity' as compared to homotopical one as in definition 7.3.3.

7.4 Homology Form of Cauchy's Theorem

Recall that while defining line integrals we first considered differentiable paths. Then, using the additivity property of the integral so obtained under subdivision of arcs, we could immediately generalize the definition of the integral over contours (which are, by definition, piecewise differentiable paths). We can now go one step further and allow our contours to have finitely many discontinuities also. (After all, recall that finitely many jump discontinuities do not cause any problem in the Riemann integration theory.) But then this is nothing but merely taking a finite number of contours γ_i together. Guided by the property that the integral over two non overlapping contours is the sum of the integrals over the two contours individually, and by the property that the integral over the inverse path is the negative of the integral, we now introduce a formal definition:

Definition 7.4.1 By a 'chain' we shall mean a finite formal sum $\sum_j n_j\gamma_j$, where n_j are any integers, and γ_j are contours. In this sum if each γ_j is a closed contour, then the sum is called a cycle. Observe that it does not hurt us if some of the integers n_j are zero. However, we do not generally write such terms in the summation. We can add two chains and rewrite the sum by 'collecting terms' whenever more than one integral is involved over the same contour. The support of a chain is the set of all image points of all those γ_j for which n_j is not zero.

Remark 7.4.1 If $\gamma = \gamma_1 \cdot \gamma_2$ then note that the two chains γ and $\gamma_1 + \gamma_2$ are different but the integrals are the same:

$$\int_\gamma f\,dz = \int_{\gamma_1} f\,dz + \int_{\gamma_2} f\,dz = \int_{\gamma_1+\gamma_2} f\,dz.$$

In anticipation of what is to come we now make another definition.

Definition 7.4.2 Two chains γ_1, γ_2 in Ω are said to be homologous to each other in Ω if for all holomorphic functions f on Ω the integrals are the same:

$$\int_{\gamma_1} f(z)dz = \int_{\gamma_2} f(z)dz$$

We express this by writing $\gamma_1 \sim \gamma_2$. A chain is said to be null homologous if it is homologous to the chain 0.

Remark 7.4.2 It is fairly easy to see that
(i) being homologous is an equivalence relation;
(ii) sum of two null homologous chains is null homologous;
(iii) two path-homotopic paths are homologous as chains (homotopy invariance theorem);
(iv) a null homologous chain is homologous to a cycle.

Following (iv) we shall also call any chain which is homologous to a cycle also a cycle. (This is not a standard terminology but introduced only temporarily only for convenience.)

The following lemma is central in the proof of Cauchy's theorem. Primarily, instead of asserting that the formula holds for all contours it says there is a special one for which it holds. Even though we could do with a little weaker version of this lemma, we have chosen this form of the lemma, which can be used for other purposes later. It has its own importance having a certain topological content.

Lemma 7.4.1 Let U be an open subset of $\mathbb{C}$ and K be a compact subset of U. Then there exists a cycle γ in $U \setminus K$ and an open subset $U' \subset U \setminus \operatorname{supp}\gamma$ such that $K \subset U'$ and for all points $z \in U'$, $\eta(\gamma, z) = 1$ and for any holomorphic function f on U' we have,

$$f(z) = \frac{1}{2\pi\imath}\int_\gamma \frac{f(\xi)}{\xi - z}\,d\xi, \quad z \in U'. \tag{7.10}$$

Proof: Since $\mathbb{C} \setminus U$ and K are disjoint closed sets and K is compact, we have,

$$\delta := d(\mathbb{C} \setminus U, K) = \inf \{|z_1 - z_2| \ : \ z_1 \in \mathbb{C} \setminus U, \ z_2 \in K\} > 0.$$

Choose $0 < \mu < \delta/3$. Raise a grid of horizontal and vertical lines with distance between consecutive parallel lines $= \mu$. Let $\mathcal{R} = \{R_j\}$ denote the collection of all little squares belonging to this grid which are at a distance $\leq \delta/3$ from K. Since K is compact, this collection has only finitely many squares. We shall denote by ∂R_j, the contour obtained by tracing the boundary of a square R_j in the counter clockwise sense. (It does not matter where you start off.)

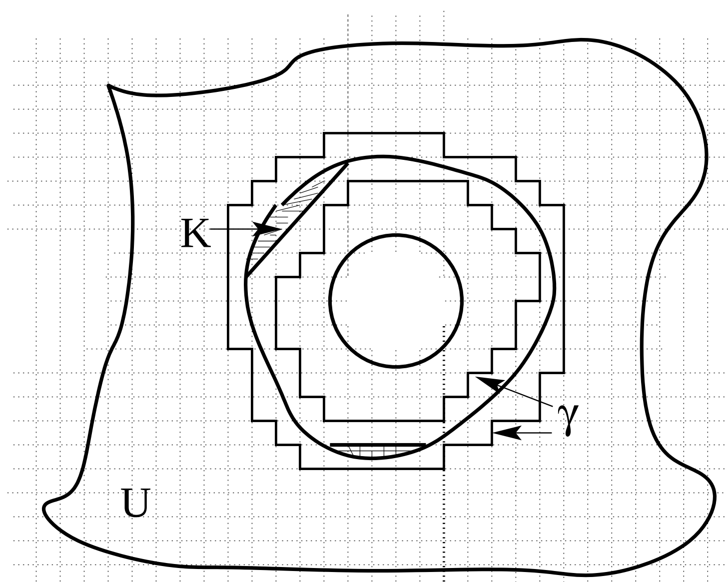


Fig. 38

Put $\gamma' := \sum_j \partial R_j$ and $R = \cup_j R_j$.

Then clearly γ' is a cycle in U and $K \subset R$. Observe that γ' is a chain consisting of directed edges of squares in the collection $\mathcal{R}$. We delete each pair of edges which are opposite of each other occurring in γ' to obtain a cycle γ. It is not very difficult to see that γ is also a cycle in U. In any case, integrals over either of these chains γ', γ will be the same for all 1-forms. (In particular γ and γ' are homologous to each other and hence we can call γ also a cycle!) Clearly the support of γ is contained in U. Moreover, $\operatorname{Supp} \gamma$ does not intersect K at all. For, if any edge intersects K, then the squares on either side of the edge are in the collection $\mathcal{R}$ and hence, the edge will occur twice, once in each direction, so gets deleted. This also shows that K is contained in the interior of R. Put $U' = \operatorname{int} R$, the interior of R.

Now, given any point z in U', since z does not lie on the support of γ, it follows that $\eta(\gamma, z)$ makes sense. If $z \in int\, R_k$, then clearly $\eta(\partial R_j, z) = 1$ if $j = k$ and $= 0$ otherwise. (See theorem 5.6.1). Therefore,

$$\eta(\gamma, z) = \eta(\gamma', z) = 1 \tag{7.11}$$

for all points $z \in \cup_k int\, R_k$. Since every point $w \in U'$ is a limit point of $\cup_k int\, R_k$ by continuity of the winding number, it follows that (7.11) is valid for all points of U'.

To see the second part, let z be in the interior of one of the R_j's. Then

$$\frac{1}{2\pi\imath}\int_{\partial R_k} \frac{f(\xi)}{\xi - z}\, d\xi = \begin{cases} f(z) & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

Therefore,

$$f(z) = \frac{1}{2\pi\imath}\int_{\gamma'} \frac{f(\xi)}{\xi - z}\, d\xi$$

for all points in $\cup_k int\, R_k$. Again, by continuity of both sides, the validity of the equation (7.10) for all points of U' follows. ♠

We are now ready to prove the equivalence of (ii) and (v) of Theorem 7.3.5.

Theorem 7.4.1 Homology Version of Cauchy's Theorem: *Let Ω be an open set in $\mathbb{C}$ and γ be a cycle in Ω. Then the following conditions on γ are equivalent:*
(i) $\displaystyle\int_\gamma f dz = 0$ *for all holomorphic functions f on Ω, i.e., γ is null homologous in Ω.*
(ii) $\eta(\gamma, a) = 0$, *for all $a \in \mathbb{C} \setminus \Omega$.*

Proof: Suppose (i) holds. For any point $a \in \mathbb{C} \setminus \Omega$, the function $\dfrac{1}{z - a}$ is holomorphic on Ω and hence from (i), we obtain

$$\eta(\gamma, a) = \frac{1}{2\pi\imath}\int_\gamma \frac{dz}{z - a} = 0.$$

This proves the statement: (i)$\Longrightarrow$ (ii).

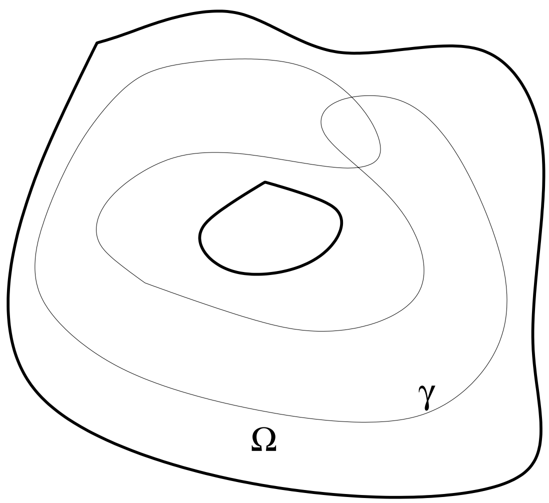


Fig. 39

To prove (ii) $\Longrightarrow$ (i), enclose the support of γ inside a disc D. Consider the closure A of the union of all components of $\mathbb{C} \setminus \text{supp}\ \gamma$ on which $\eta(\gamma, z) \neq 0$. Then $A \subset D$ and hence bounded. Moreover (ii) implies that $A \subset \Omega$. We can now apply lemma 7.4.1, to the situation $K = A \cup \text{supp}\,\gamma \subset \Omega := U$ to obtain

(i) a cycle ω in $\Omega \setminus K$ and

(ii) an open set U' such that $K \subset U' \subset \Omega \setminus \text{supp}\ \omega$ with the property that for all points $z \in U'$, $\quad \eta(\omega, z) = 1$ and for all holomorphic function f on U', we have,

$$f(z) = \frac{1}{2\pi\imath} \int_\omega \frac{f(\xi)}{\xi - z}\, d\xi, \quad z \in U'.$$

On the other hand, since supp $\omega \cap A = \emptyset$, it follows that $\eta(\gamma, \xi) = 0$ for all $\xi \in$ supp ω. Therefore,

$$\begin{aligned} \int_\gamma f(z)\, dz &= \int_\gamma \left(\frac{1}{2\pi\imath} \int_\omega \frac{f(\xi)}{\xi - z}\, d\xi \right) dz \\ &= \int_\omega \left(\frac{1}{2\pi\imath} \int_\gamma \frac{dz}{\xi - z} \right) f(\xi)\, d\xi = \int_\omega \eta(\gamma, \xi) f(\xi)\, d\xi = 0. \end{aligned}$$

(For justification in the change in the order of integration see (4.12). This completes the proof of the theorem. ♠

Remark 7.4.3 We have remarked earlier that condition (ii) of the above theorem may be referred to as *homologically simple connectivity.* Often elementary books in complex analysis call this property itself simple connectivity. This is justified so far as we are talking about domains in $\mathbb{C}$ but not in general. Also note that this comes very close to the geometric simple connectivity that we have introduced in definition 4.4.1. The following theorem gives another purely set-topological characterization of this property. We may call this one, the topological simple connectivity.

Theorem 7.4.2 *Let Ω be a domain is $\mathbb{C}$. Then the following conditions on Ω are equivalent.*
(i) $\eta(\gamma, z) = 0$ *for every cycle* γ *in* Ω *and every point* $z \in \mathbb{C} \setminus \Omega$.
(ii) $\widehat{\mathbb{C}} \setminus \Omega$ *is connected.*

Proof: To prove (i) $\Longrightarrow$ (ii), assume $\widehat{\mathbb{C}} \setminus \Omega$ is not connected. Write $\widehat{\mathbb{C}} \setminus \Omega = X \coprod Y$, as a disjoint union of non empty closed subsets. In particular, both X and Y are compact. Say, $\infty \in Y$. Then $X \subset \mathbb{C}$ is compact. Take $U = X \cup \Omega = \widehat{\mathbb{C}} \setminus Y$. Then U is open. So, we can apply lemma 7.4.1 to conclude that there is a cycle γ in Ω such that $\eta(\gamma, a) = 1$ for all $a \in X$ which is a contradiction to (i).

To prove (ii) $\Longrightarrow$ (i), we may as well assume that γ is a contour. Recall from remark 5.6.1.7, that $\eta(\gamma, -)$ is a continuous integer valued function on $\widehat{\mathbb{C}} \setminus Im(\gamma)$ for any contour γ. Since $\widehat{\mathbb{C}} \setminus \Omega$ is a connected subset of $\widehat{\mathbb{C}} \setminus supp(\gamma)$, $\eta(\gamma, -)$ is a constant equal to $\eta(\gamma, \infty) = 0$. ♠

Remark 7.4.4 One would love to have a characterization of simple connectivity of a domain in $\mathbb{C}$ without bringing in the point at infinity. Intuitively suggested is the following condition:
(iii) $\mathbb{C} \setminus \Omega$ has **no** bounded (and hence, compact) components.

It is not hard to see that (iii) implies (ii) (exercise). However, the proof of (ii) or (i) implies (iii) involves deeper point-set-topology. (For a complete proof see R. Narasimhan's book [N].) Thus, we could add two more statements to the list of statments in theorem 7.3.5, which are all equivalent to each other.

Exercise 7.4

1. Prove the following form of theorem 5.7.1.

   Homology Form of Residue Theorem : Let Ω be an open set in $\mathbb{C}$, S be a discrete subset of Ω, $\Omega' = \Omega \setminus S$. Let γ be a cycle in Ω' such that $\eta(\gamma, a) = 0$ for all $a \in \mathbb{C} \setminus \Omega$. Then for any holomorphic function f on Ω', we have

$$\frac{1}{2\pi\imath} \int_\gamma f dz = \sum_{a \in S} \eta(\gamma, a) R_a(f). \tag{7.12}$$

2. Prove the following **Generalized Logarithmic Residue Theorem:** Let g be a holomorphic function and f be a meromorphic function which is not identically 0 in a domain Ω. Let $a_1, \ldots, a_k, \ldots$ and $b_1, \ldots, b_l, \ldots$ be a listing of the zeros and poles of f, with each zero and pole being repeated as many times as its order in the listing. Then for any closed contour γ in Ω which does not pass through any of the a_j, b_k's, and such that $\eta(\gamma, z) = 0$ for all $z \in \mathbb{C} \setminus \Omega$, we have,
$$\frac{1}{2\pi\imath}\int_\gamma \frac{g(z)f'(z)}{f(z)}dz = \sum_j \eta(\gamma, a_j)g(a_j) - \sum_k \eta(\gamma, b_k)g(b_k). \tag{7.13}$$

3. Formulate a similar generalization (as in the above exercises) of theorem 5.7.4 and prove it.

4. Let Ω be a (homologically) simply connected domain and $A = \{a_1, \ldots, a_m\} \subset \Omega$. Choose discs D_j with center at a_j and contained in Ω so that $D_j \cap D_k = \emptyset,\ j \neq k$. Show that every cycle in $\Omega \setminus A$ is homologous to a cycle of the form $\sum_j n_j \partial D_j$. Thus, you may think of the contours ∂D_j forming a generating set for homology of $\Omega \setminus A$. For this reason, ∂D_j are called **fundamental cycles**. For any differential form ω on $\Omega \setminus A$, the values $\int_{\partial D_j} \omega$ are called its **fundamental periods**.

5. Let Ω and A be as in the previous exercise. Suppose ω is a differential 1-form on $\Omega \setminus A$ such that $\int_{\partial D_j} \omega = 0$ for $1 \le j \le m$. Show that ω is exact.

7.5 Miscellaneous Exercises to Ch. 7

1. Show that there cannot be any rational mapping defining a biholomorphic mapping of $\boldsymbol{H}$ onto the upper part of the parabola $y = x^2$. [Hint: Use exercise 15 in Misc. Ex to ch. 3 and exercise 7.1.5.]

2. Let f be an entire function. Show that the Maclaurin's series for f converges uniformly to f on $\mathbb{C}$ iff f is a polynomial.

3. Let f be a holomorphic function on $\Omega := B_r(w) \setminus \{w\}$. Suppose further that one of the following conditions holds:
(i) $\Re(f)$ is bounded on Ω.
(ii) $\Im(f)$ is bounded on Ω.
(iii) There exists $z_0 \neq 0$, & $R > 0$ such that $f(\Omega) \cap \{tz_0 \ :\ t > R\} = \emptyset$.
Show that w is a removable singularity of f.

4. Show that for any $r > 0$ and $0 < r_1 < r_2$, there is no biholomorphic mapping of the punctured disc $A(0; 0, r)$ and the annulus $A(0; r_1, r_2)$.

5. Suppose $f : \bar{\mathbb{D}} \to \mathbb{C}^*$ is a continuous function which holomorphic in the interior. If $|f(z)| = 1$ for all $|z| = 1$ then show that f is a constant.

6. Let $p(z) = \sum_0^n a_k z^k$ be such that $r^j |a_j| > \sum_{k \neq j} |a_k| r^k$ for some $r > 0$ and $a_j \neq 0$. Then show that $p(z)$ has exactly j roots inside $B_r(0)$. Use this to determine the number of zeros of the following polynomials inside the respective annuli:
   (a) $p_1(z) = z^4 - 6z + 3$; $15/32 < r < 1$ and $1 < r < 63/32$.
   (b) $p_2(z) = 9z^5 + 5z - 3$; $0 < r < 1/2$ and $1/2 < r < 1$.
   (c) $p_3(z) = z^8 - z^7 - 4z^2 - 1$; $0 < r < 1$ and $1 < r < 2$.

7. Given $\lambda > 1$, show that there is precisely one root of the equation $ze^{\lambda - z} = 1$ inside $\mathbb{D}$.

8. Suppose $f(z) = \sum a_k (z - z_0)^k$ is analytic in a disc $B_r(z_0)$. Let $0 < s < r$ be such that f has no zeros on $\partial B_s(0)$. Then show that for sufficiently large n, the polynomial $\sum_{k=0}^n a_k (z - z_0)^k$ has exactly as many zeros as f inside $B_s(z_0)$.

9. Let $p(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n$. Suppose all the roots of p are inside the circle $|z| = R$. For $1 \leq k \leq n$, put $\alpha_k := \dfrac{1}{2\pi\imath} \displaystyle\int_{|z|=R} \frac{z^k p'(z)}{p(z)} dz$. Show that $\alpha_1 = -a_1$. More generally show that

$$\alpha_k + a_1 \alpha_{k-1} + \cdots + a_{k-1} \alpha_1 + k a_k = 0.$$

10. Show that for $n \geq 2$ and any complex number w, the equation $z^n + w(z + 1) = 0$ has at least one root in the closed disc $|z| \leq 2$. [Hint: consider the cases according as $|w| \leq 2^n$ or not.]

11. Find the Laurent series representation of $\ln \left(\dfrac{z^r}{z^r - 1} \right)$, in the annulus $|z| > 1$ for any positive integer r.

12. If $\sum_0^\infty a_n z^n$ is a power series with radius of convergence r, what is the domain of convergence for the Laurent series $\sum_{-\infty}^\infty a_{|n|} z^n$?

13. Let $0 < r_1 < r_2 < \infty$ and let f be a holomorphic function on $A := A(0; r_1, r_2)$. Suppose that for every sequence $\{z_n\}$ in A such that $|z_n| \longrightarrow r_j, j = 1, 2$, we have,

$\lim_n f(z_n) = 0$. Show that $f \equiv 0$. Can we take $r_1 = 0$ or $r_2 = \infty$ and still conclude the same thing?

14. Let f be holomorphic except at finitely many isolated singularities. Then for any positive integer n, show that the sum of the residues of f is equal to the sum of the residues of ϕ_n, where $\phi_n(z) = z^{n-1}f(z^n)$. If p is a polynomial of degree n what can you say about the residues of f and that of the function defined by $h(z) = p'(z)f(p(z))$?

15. (For this exercise, assume the fact that the image of any contour never contains a non empty open set in $\mathbb{C}$.) Let $f : \Omega \to \mathbb{C}$ be a non constant holomorphic function and γ be a cycle in Ω such that $A = \{z \in \Omega \ : \ \eta(\gamma, z) \neq 0\}$ is non empty. Show that $f(\gamma)$ separates $\mathbb{C}$.

16. Let f be holomorphic in a domain Ω and γ be a cycle in Ω such that $A = \{z \in \Omega \ : \ \eta(\gamma, z) \neq 0\}$ is non empty. Suppose f maps each component of the support of γ inside a vertical line in $\mathbb{C}$. Show that f is a constant.

17. Let γ be a simple closed curve consisting of line segments parallel to the x- axis or the y-axis (i.e., an axial contour). Establish Jordan curve theorem in this case, viz., show that $\mathbb{C} \setminus \gamma$ has precisely two components.

18. Given any two disjoint subsets K_1 and K_2 of $\mathbb{C}$, such that K_1 compact and K_2 closed, show that there exists an axial, simple closed contour which separates K_1 from K_2.

Chapter 8

Convergence in Function Theory

8.1 Sequences of Holomorphic Functions

One of the most important results in the theory of convergence of functions is the majorant criterion of Weierstrass(see theorem 1.4.8). Following this, we make some formal definitions: Let $f_n : X \longrightarrow \mathbb{C},\ n \geq 1$ be a sequence of functions.

Definition 8.1.1 We say the series $\sum_n f_n$ is *compactly convergent* in X if restricted to any compact subset of X, it is uniformly convergent.

Definition 8.1.2 Let us introduce the notation

$$|f|_A = \text{Sup}\,\{|f(z)| \ : \ z \in A\}$$

A series $\sum f_n$ is said to be *normally convergent*[1] in X if for every point $x \in X$, there exists a neighborhood U such that $\sum_n |f_n|_U < \infty$.

Observe how Weierstrass's criterion has been adopted into a definition here: for a normally convergent series, the terms $|f_n|$ play the role of majorants, in the neighborhood U. It follows that every normally convergent series is locally uniformly convergent in X. Indeed, for series of continuous functions over domains in euclidean spaces(local compactness!), normal convergence is a convenient terminology for absolute local uniform convergence. For a normally convergent series $\sum_n f_n$ of continuous functions, it follows that the sum $\sum f_n$ is also continuous. Linear combination of normally convergent series

[1] Introduced by René Baire(1874-1932) apologetically, this concept is often encountered in functional analysis wherein the spaces on which functions are being studied need not be locally compact.

is normally convergent. Also the same is true of Cauchy products of normally convergent series. In addition, because of the built-in absolute convergence, we have the following two results.

Theorem 8.1.1 *Every subseries of a normally convergent series is normally convergent.*

Theorem 8.1.2 Rearrangement Theorem: *Let $\sum f_n$ be a normally convergent series. Then for any bijection $\tau : \mathbb{N} \longrightarrow \mathbb{N}$ the series $\sum f_{\tau(n)}$ is also normally convergent to the same sum.*

Proof: Let f be the sum $\sum f_n$. Let $x \in X$ and U be a neighborhood of x such that $\sum |f_n|_U < \infty$. By the rearrangement theorem for absolute convergent series of complex or (real) numbers, it follows that $\sum |f_{\tau(n)}|_U$ is convergent to $\sum_n |f_n|_U$. This means that $\sum f_{\tau(n)}$ is normally convergent. ♠

The following theorem due to Weierstrass, guarantees the holomorphicity of the limit under normal convergence. It also provides us the validity of term-by-term differentiation.

Theorem 8.1.3 Weierstrass's Convergence Theorem: *Suppose that we are given a sequence f_n of compactly convergent holomorphic functions in a domain Ω. Then the limit function f is holomorphic and the sequence $f_n^{(k)}$ compactly converges to $f^{(k)}$ on Ω for every positive integer k.*

Proof: Observe that the limit function f is continuous in Ω by theorem 1.4.9. Let γ be the boundary of a triangle contained in Ω. Then the sequence f_n converges uniformly to f on γ. Hence it follows that

$$\lim_n \left(\int_\gamma f_n(z)dz \right) = \int_\gamma f(z)dz.$$

By Cauchy-Goursat theorem, each term on the lhs vanishes and hence rhs also vanishes. By Morera's theorem, it follows that f is holomorphic in Ω.

For the second part, to each closed ball $L = B_{2r}(z_0)$ contained in Ω, put $K = B_r(z_0)$. First, we use Cauchy's integral formula to find M_r such that $|f_n^{(k)} - f^{(k)}|_K \leq M_r |f_n - f|_L$ for all n, as follows:

$$|f_n^{(k)}(z) - f^{(k)}(z)| = \left| \frac{k!}{2\pi\imath} \int_C \frac{f_n(\xi) - f(\xi)}{(\xi - z)^{k+1}} d\xi \right| = \frac{k!}{2\pi} . \frac{|f_n - f|_L}{r^{k+1}} .4\pi r,$$

where $C = \partial B_{2r}(z_0)$ $z \in K$. So, we take $M_r = 2\frac{k!}{r^k}$. Since f_n uniformly converges to f on L, it follows that so does $f_n^{(k)}$ to $f^{(k)}$ on K. ♠

Example 8.1.1 As a typical example, consider the series $\sum_{n=1}^{\infty} \frac{1}{n^z}$. If $\Re(z) > 1 + \epsilon$, then $|n^z| = n^{\Re(z)} > n^{1+\epsilon}$ and hence the series $\sum_{1}^{\infty} \frac{1}{n^{1+\epsilon}}$ is a majorant for the given series. This implies that $\sum_{n=1}^{\infty} \frac{1}{n^z}$ is a normally convergent series and hence defines a holomorphic function $\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$ in the right-half plane $\mathbb{G}_1 = \{z : \Re(z) > 1\}$. In fact, it is uniformly convergent in $\mathbb{G}_{1+\epsilon} = \{x + \imath y : x > 1 + \epsilon\}$. This function is called Riemann's zeta-function. We shall come back to this again in a later section.

Remark 8.1.1 In the above theorem, one can start with a series which is normally convergent in Ω and then conclude similarly with normal convergence of the derived series. However, you may have learnt that a statement similar to that of the above theorem in the real case is false. A typical counter example is

$$f_n(x) = \frac{x}{1 + nx^2}$$

defined on the interval $(-1, 1)$. It is not difficult to show that $\{f_n\}$ compactly converges to the function f which is identically zero on the interval. But, $0 = f'(0) \neq \lim_{n\to\infty} f_n'(0) = 1$. (Why then does the sequence $f_n(z) = \frac{z}{1 + nz^2}$ not provide a counter example to the above theorem?) We shall see more interesting examples in section 8.6.

Theorem 8.1.4 Weierstrass' Double Series Theorem. *Let $f_m(z) = \sum_n a_{mn} z^n$ be a sequence of series convergent in a disc B for all $m \in \mathbb{N}$. Suppose the series $f(z) = \sum f_m(z)$ converges normally in B. Then for each n the series $b_n = \sum_m a_{mn}$ is convergent and f is represented in B by the convergent power series $f(z) = \sum_n b_n z^n$.*

Proof: Clearly, by the above theorem, f is holomorphic in B. Also $f^{(n)} = \sum_m f_m^{(n)}$ for all n. Moreover f is represented by the Taylor's series, $f(z) = \sum_n f^{(n)}(0) z^n / n!$. By substituting expressions for $f^{(n)}(0)$, we obtain the result. ♠

Remark 8.1.2 As an illustration that the Cauchy product of two compactly convergent series need not be convergent, consider for each n, $f_n = g_n = (-1)^n / \sqrt{n+1}$, the constant function for all n. Clearly $\sum_n f_n = \sum_n g_n$ are convergent. But the modulus of the k^{th} term of the Cauchy product satisfies:

$$|h_k| = \sum_{m=0}^{k} [(m+1)(k-m+1)]^{-1/2} > \sum_{m=0}^{k} \frac{1}{k+1} = 1.$$

Hence the Cauchy product, $\sum h_k$ is not convergent. This also gives an example of a compactly convergent series which is not normally convergent.

We shall end this section with a celebrated result due to Hurwitz[2] concerning preservation of the zeros under compact convergence.

Theorem 8.1.5 Hurwitz: *Let f_n be a sequence of holomorphic functions compactly convergent to f in a region Ω. If f_n have no zeros in Ω then either $f \equiv 0$ or f has no zeros in Ω.*

Proof: Assuming that $f \not\equiv 0$, since the zeros of f are isolated, given $z_0 \in \Omega$, we can choose $r > 0$ such that $f(z) \neq 0$ on $\partial B_r(z_0) := C$. Then $1/f_n(z)$ converges uniformly to $1/f(z)$ on C and since $f_n'(z)$ converges uniformly to $f'(z)$ on C, it follows that $f_n'(z)/f_n(z)$ converges uniformly to $f'(z)/f(z)$ on C. Hence, we have,

$$\int_C \frac{f'(z)}{f(z)} dz = \lim_n \int_C \frac{f_n'(z)}{f_n(z)} dz.$$

Since f_n have no zeros in Ω, every term on rhs is zero. On the other hand, the lhs counts the number of zeros of f inside C. Hence in particular, it follows that $f(z_0) \neq 0$. ♠

Corollary 8.1.1 In the situation of the above theorem, assume further that f_n is injective in Ω for all n sufficiently large. Then the limit function f is either a constant or injective.

Proof: Assume that the limit function f is not a constant. Given $z_0 \in \Omega$, consider the sequence $g_n : \Omega' := \Omega \setminus \{z_0\} \to \mathbb{C}$ defined by

$$g_n(z) = f_n(z) - f_n(z_0).$$

Since $\{f_n\}$ are injective, it follows that $\{g_n\}$ have no zeros in Ω'. Therefore the limit function g is either identically zero or has no zeros in Ω'. The first case is ruled out because that would imply $f(z) = f(z_0)$, for all $z \in \Omega$. Therefore, g has no zeros in Ω'. This means $f(z) \neq f(z_0), z \neq z_0$. Since z_0 is arbitrary, this implies f is injective. ♠

For sharper results in this direction, see the excellent book of Remmert p. 261-262.

Exercise 8.1

[2]Adolf Hurwitz from Zurich is known for his work on analytic functions and Cantor's set theory. He should not be confused with W. Hurewicz, who is the author of a famous book on dimension theory jointly with Wallman.

1. Let $f_n(z) = 1 + z + \dfrac{z^2}{2!} + \cdots + \dfrac{z^n}{n!}$, $Z_n := \{z \ : \ f_n(z) = 0\}$ and let $\tau_n = d(0, Z_n)$, be the distance between 0 and the set Z_n. Show that $\tau_n \longrightarrow \infty$ as $n \longrightarrow \infty$.

2. Show that the series $\sum_{n\geq 1} \frac{(-1)^n}{z+n}$ is compactly convergent but not normally convergent in $\mathbb{C} \setminus \{-1, -2, -3, \ldots, \}$.

8.2 Convergence Theory for Meromorphic Functions

Let us first recall a definition.

Definition 8.2.1 A function f is called *meromorphic* in a domain Ω if there is a closed discrete subset $P(f)$ of Ω such that f is holomorphic in $\Omega \setminus P(f)$ and has poles at each point of $P(f)$. Naturally the set $P(f)$ is called the *pole-set* of f.

It follows that $P(f)$ is always countable, since every discrete subset of $\mathbb{C}$ is. (See exercise 5.1.2.) Observe that, in the definition of a meromorphic function f, we allow the set $P(f)$ to be empty and thus all holomorphic functions on Ω are also meromorphic. Since we know that a meromorphic function tends to ∞ at its poles, we can think of them as continuous functions on Ω with values in the extended complex plane $\widehat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ by mapping each pole onto ∞. Recall that if $z = z_0$ is a pole of f of order $n \geq 1$, then it is a pole of order $n+1$ for f'. Thus if f is meromorphic, so is f' and $P(f) = P(f')$. Let us denote the set of all meromorphic functions on Ω by $\mathcal{M}(\Omega)$ and the set of holomorphic functions on Ω by $\mathcal{H}(\Omega)$. Observe that for each $f \in \mathcal{M}(\Omega)$ which is not identically zero, we have $1/f \in \mathcal{M}(\Omega)$. Thus for any two $f, g \in \mathcal{H}(\Omega)$, $f + g \in \mathcal{H}(\Omega)$ and if with $g \neq 0$, then $f/g \in \mathcal{M}(\Omega)$. We laos know that if $fg \equiv 0$ then either $f \equiv 0$ or $g \equiv 0$. This makes $\mathcal{H}(\Omega)$ into an integral domain. We shall prove soon that every element $h \in \mathcal{M}(\Omega)$ can be expressed as $h = f/g$, as above, as a consequence of Weierstrass' theorem. [In algebraic terminology, this means $\mathcal{M}(\Omega)$ is the quotient field of the integral domain $\mathcal{H}(\Omega)$.]

What we are interested in doing in this section is to develop the theory of convergence for sequences of meromorphic functions as a tool to obtain many more examples of meromorphic functions. The first difficulty was that a meromorphic function is not 'a function' in the set-theoretic sense. This difficulty is immediately removed by adopting ∞ as a genuine value and it seems that our problem is resolved. Wait a minute.

Consider the sequence in $\mathcal{M}(\mathbb{C})$

$$f_n = \frac{1}{n!z^n}, \ z \neq 0, \ n \geq 1.$$

This sequence normally converges to the function τ which takes the constant value 0 except at the point 0, where it takes the value ∞. Clearly τ is not even a continuous function. The corresponding series $\sum_n f_n$ is even worse: It converges normally to $e^{1/z}$ which has an essential singularity at 0.

The situation is not at all new. We know that a sequence of continuous functions may point-wise converge to a function f which may not be continuous. This problem was resolved by strengthening the definition of convergence to that of uniform convergence. But once again, when we were dealing with smooth functions, even uniform convergence failed to produce smooth functions. However, we did not change the definition of convergence but put extra condition on the limit itself.

What should be the right thing to do here? Presently, since we are interested in producing more and more meromorphic functions, we prefer to change the definition slightly, which will enable us to talk about convergence of meromorphic functions without any hindrance. Such a condition is indicated by the above example: *we should not allow infinitely many terms in the series to have a common pole.*

Definition 8.2.2 A series $\sum f_n$ of meromorphic functions in Ω is called *compactly convergent* in Ω if for every compact subset K of Ω there exists a number $m(K)$ such that
(MF1) $n \geq m(K) \Longrightarrow P(f_n) \cap K = \emptyset$ and
(MF2) the series $\displaystyle\sum_{n \geq m(K)} f_n$ converges uniformly on K.

We say that the series $\sum_n f_n$ of meromorphic functions is *normally convergent* if for every compact subset K of Ω condition (MF1) holds and in place of (MF2) we have the stronger condition:
$(MF2')$ $\displaystyle\sum_{n \geq m(K)} |f_n|_K$ is convergent.

Condition (MF1) is called *pole dispersion condition.* Under this condition, the remaining terms in the series are pole free on K and hence in particular, continuous. Observe that, (MF1) implies that $\cup_n P(f_n)$ is discrete and closed in Ω. It is also clear that it is enough to demand that (MF1) holds for all closed discs inside Ω instead of for all compact subsets of Ω. The theorem below follows directly from the above definition.

Theorem 8.2.1 *Let $\sum f_n$ be a series of meromorphic functions in Ω compactly (resp. normally) convergent in Ω. Then there exists a unique meromorphic function f on Ω with the following property:*

For each open subset U of Ω and for each integer m such that $P(f_n) \cap U = \emptyset$, $\forall\, n \geq m$, the series $\sum_{n\geq m} f_n$ converges compactly (resp. normally) in U to a holomorphic function F_U on U such that

$$f|_U = f_0|_U + f_1|_U + \cdots + f_{m-1}|_U + F_U. \tag{8.1}$$

In particular, f is holomorphic in $\Omega \setminus \cup_n P(f_n)$.

Proof: Let U be an open subset of Ω whose closure is compact. Let $m(K)$ be as in (1) for $K = \overline{U}$. It follows that $F_U = \sum_{n\geq m(\overline{U})} f_n$, is a holomorphic function on U. Thus (8.1) defines a meromorphic function on U. If V is another open set with compact closure and the number $m(\overline{V})$ is chosen similarly,

$$f|_V = f_0|_V + f_1|_V + \cdots + f_{m(\overline{V})-1}|_V + F_V.$$

For definiteness, say, $m(\overline{U}) \leq m(\overline{V})$. Then on $U \cap V$ we have,

$$F_U = f_{m(\overline{U})} + f_{m(\overline{U})+1} + \cdots + f_{m(\overline{V})-1} + F_V$$

and hence, if follows that $f_U = f_V$. Since Ω can be covered by a family of open sets $\{U_\alpha\}$ with their closure compact, we may define $f : \Omega \longrightarrow \mathbb{C}$ by $f(z) = f_{U_\alpha}(z), z \in U_\alpha$. The rest of the claims of the theorem are easily verified. ♠

We call f the sum of the series and write $f = \sum f_n$. We emphasize that whenever you come across with an infinite sum of meromorphic functions, you should remember the pole dispersion condition (MF1). It is not at all difficult to see that linear combinations of compactly (resp. normally) convergent series of meromorphic functions produce compactly (resp. normally) convergent series again. For normally convergent series, we even have the rearrangement theorem. Perhaps not so obvious is:

Theorem 8.2.2 Term-wise Differentiation: *Let $f_n \in \mathcal{M}(\Omega)$ and let $\sum f_n = f$ be compactly (resp. normally) convergent. Then the term-wise differentiated series $\sum f_n'$ compactly (resp. normally) converges to f'.*

Proof: Let U be an open disc such that the closed disc $\overline{U} \subset \Omega$. Choose m such that $P(f_n) \cap U = \emptyset$, for all $n \geq m$. Then $\sum_{n\geq m} f_n$ converges compactly (resp. normally) to a holomorphic function F in U such that (8.1) holds. We can apply the term-wise differentiation to this partial series and get $F' = \sum_{n\geq m} f_n'$ on U. Addition of first $m-1$

terms does not violate the nature of convergence. Thus, $\sum_n f'_n$ is compactly (resp. normally) convergent in Ω. Also because of (8.1) its sum g satisfies,

$$g|_U = f'_0|_U + \cdots + f'_{m-1}|_U + F' = (f|_U)'.$$

Since this is true for all such discs in Ω, we obtain, $g = f'$, in Ω. ♠

8.3 Partial Fraction Development of $\pi \cot \pi z$.

We would like to employ the above theoretical discussion to a practical situation. The theme is somewhat in the reverse order. *We begin with a meromorphic function and try to write it as a sum of the most simple meromorphic functions.* For the sake of simplicity, we shall consider the case $\Omega = \mathbb{C}$ only. Consider for instance the case when f has finitely many poles $b_1, \ldots, b_k$ with its respective singular parts $P_j\left(\frac{1}{z-b_j}\right)$. We know that P_j are actually polynomials and we have $f(z) = \sum_j P_j\left(\frac{1}{z-b_j}\right) + g(z)$, where g is an entire function. This kind of representation has many advantages as it directly gives much information about the function. In the general case when f has infinitely many poles the sum $\sum_j P_j\left(\frac{1}{z-b_j}\right)$ may not be convergent in any sense. For instance, if we take $P = \mathbb{N}$ and $P_j(z) = z,\ \forall\, j \in \mathbb{N}$, then we get the sum $\sum_{n\geq 0} \frac{1}{z-n}$ which does not converge. Let us solve this problem for the function $\pi \cot \pi z$ in a completely ad hoc fashion.

Introducing the Eisenstein's summation notation

$$\sum_{-\infty}^{\infty}{}_e := \lim_{n\longrightarrow\infty} \sum_{-n}^{n}, \tag{8.2}$$

we write

$$E_1(z) = \sum_{-\infty}^{\infty}{}_e \frac{1}{z+n}. \tag{8.3}$$

We would like to draw the attention of the reader to the difference in the meanings of the two notions viz., $\sum_{-\infty}^{\infty}$ and $\sum_{-\infty}^{\infty}{}_e$. These are similar to the improper integrals $\int_{-\infty}^{\infty}$ and the Cauchy's principle value, $PV\int_{-\infty}^{\infty}$, respectively. For any sequence $\{f_n\}_{n\in\mathbb{Z}}$ we define

$$s_n = \sum_{k=-n}^{n} f_k, \text{ and } \sum_{-\infty}^{\infty}{}^{e} f_n = \lim_{n\to\infty} s_n.$$

Thus (8.3) can be rewritten as

$$E_1(z) := \lim_{n\longrightarrow\infty} \sum_{-n}^{n} \frac{1}{z+k} = \frac{1}{z} + \sum_{1}^{\infty} \frac{2z}{z^2-n^2}. \tag{8.4}$$

We shall first show that (8.4) converges normally in $\mathbb{C}$. For any $r > 0$, we have,

$$|z \pm n|^k \geq (n-r)^k, \quad k \geq 1, \quad |z| \leq r < n.$$

Hence, on the disc $K = B_r(0)$, we have, $|z^2 - n^2| > n^2 - r^2$. Therefore,

$$\left| \frac{2z}{(z^2-n^2)^k} \right|_K \leq \frac{2r}{(n^2-r^2)^k} \quad \text{for } k \geq 1, n > r.$$

Since $\displaystyle\sum_{n>r} \frac{2r}{(n^2-r^2)^k}$ is convergent, the normal convergence of the series above follows. Thus E_1 is a meromorphic function in $\mathbb{C}$. In order to show that this function is nothing but $\pi cot\, \pi z$, we first observe that:

(OD1) $\pi \cot \pi z$ is an odd function with its pole set equal to $\mathbb{Z}$ and its principle parts equal to $\frac{1}{z-n}$ and

(OD2) $\pi \cot \pi z$ satisfies the *duplication formula* (*double angle formula*)

$$\boxed{2(\pi\cot(2\pi z)) = \pi\cot\pi z + \pi\cot\pi\left(z + \frac{1}{2}\right).} \tag{8.5}$$

These two properties are easily verified. We next show that:

Lemma 8.3.1 Any meromorphic function which satisfies conditions (OD1) and (OD2) has to be equal to $\pi \cot \pi z$.

Proof: Let f be a meromorphic function on $\mathbb{C}$ with the said properties and consider $G(z) = f(z) - \pi \cot \pi z$. Then G is an entire function which is odd and satisfies

$$\begin{aligned} 2G(2z) &= 2f(2z) - 2\pi\cot 2\pi z \\ &= f(z) + f\left(z+\frac{1}{2}\right) - \pi\cot\pi z - \pi\cot\pi\left(z+\frac{1}{2}\right) \\ &= G(z) + G\left(z+\frac{1}{2}\right). \end{aligned} \tag{8.6}$$

Also, observe that $G(0) = 0$ since G is an odd function. Suppose G is not identically zero. Then by the maximum principle, it follows that there exists $z_0 \in \partial B_2(0)$ such that $|G(z)| < |G(z_0)|$ for all $z \in B_2(0)$. Now observe that $z_0/2, (z_0+1)/2$ are both in $B_2(0)$. Therefore, using (8.6) above, we get,

$$|2G(z_0)| = \left| G\left(\frac{z_0}{2}\right) + G\left(\frac{z_0+1}{2}\right)\right| \le \left|G\left(\frac{z_0}{2}\right)\right| + \left|G\left(\frac{z_0+1}{2}\right)\right| < 2|G(z_0)|$$

which is absurd. Therefore G is identically zero, as claimed.

Finally, we verify that the function E_1 satisfies the conditions (OD1) and (OD2) above: That E_1 is meromorphic with pole set $\mathbb{Z}$ is clear from the earlier analysis. That its principal part is equal to $1/(z-n)$ at $z = n$ also follows from this analysis. The oddness is checked directly by the definition. To see that E_1 satisfies the duplication formula, we first consider the partial sum

$$s_n(z) = \frac{1}{z} + \sum_{k=1}^{n}\left(\frac{1}{z+k} + \frac{1}{z-k}\right).$$

We directly verify that $s_n(z) + s_n(z+1/2) = 2s_{2n}(z) + \dfrac{2}{2z+2n+1}$. Taking the limits now yields the required result. ♠

Thus we obtain,

$$\boxed{\pi \cot \pi z = \frac{1}{z} + \sum_{1}^{\infty} \frac{2z}{z^2 - n^2}.} \tag{8.7}$$

Exercise 8.3

1. Show that

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{-\infty}^{\infty} \frac{1}{(z-n)^2}; \qquad \pi^3 \frac{\cot \pi z}{\sin^2 \pi z} = \sum_{-\infty}^{\infty} \frac{1}{(z+n)^3}.$$

   [Hint: Differentiate E_1 once and then again.] Justify your steps.

2. Here is an amusing way of computing $\displaystyle\int_0^\infty \frac{\sin x}{x} dx$ without using residues: First, obtain the partial fraction development

$$\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{1}^{\infty} \frac{(-1)^n 2z}{z^2 - n^2} = \sum_{-\infty}^{\infty}{}_{e} \frac{(-1)^n}{z+n}.$$

   [Hint: Begin with the identity

$$\frac{\pi}{\sin \pi z} = \frac{\pi}{2} \cot \frac{\pi z}{2} - \frac{\pi}{2} \cot \frac{\pi(z-1)}{2}$$

and use (8.7).] Now, replace z by $\pi^{-1}z$ and use the fact $\sin(n\pi + z) = (-1)^n \sin z$ to rewrite the above formula:

$$1 = \sum_{-\infty}^{\infty}{}^{e} \frac{\sin(n\pi + z)}{n\pi + z}. \tag{8.8}$$

Now integrate this on the interval $[0, \pi]$ and simplify.] Justify all the steps.

3. Obtain the Taylor coefficients of $\dfrac{1}{z^2 - n^2}$ by using geometric series expansion. Next using Weierstrass' double series theorem, get the Taylor coefficients of the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{z^2 - n^2}$ around 0. (Use the fact that the function is even.) Use this to obtain the Taylor expansion of $E_1(z) - \dfrac{1}{z}$ around 0 :

$$E_1(z) - \frac{1}{z} = -\sum_{1}^{\infty} 2\zeta(2n) z^{2n-1}.$$

Comparing with the Taylor series for $\pi \cot \pi z - \dfrac{1}{z}$, obtain the celebrated **Euler's formula** for the zeta-function:

$$\boxed{\zeta(2n) = (-1)^{n-1} \frac{(2\pi)^{2n}}{2(2n)!} B_{2n}, \quad n = 1, 2, \ldots} \tag{8.9}$$

Here B_{2n} denote the Bernoulli numbers; see exercises 4.6.3. By computing the few Bernoulli numbers obtain the specific Euler's formulae:

$$\sum \frac{1}{n^2} = \frac{\pi^2}{6}; \qquad \sum \frac{1}{n^4} = \frac{\pi^4}{90}; \qquad \sum \frac{1}{n^6} = \frac{\pi^6}{945}, \cdots$$

4. Using the partial fraction development for $\dfrac{\pi}{\sin \pi z}$, show that

$$\frac{\pi}{4} = \sum_{1}^{\infty} \frac{(-1)^{n-1}}{2n - 1}.$$

8.4 Mittag-Leffler Theorem

The method employed in the previous section to arrive at a series representation for $\pi \cot \pi z$, we simply combined pair-wise singular parts at two different poles of the function to arrive at a convergent series. This method is completely ad hoc and demands a fair amount of symmetry so as to be effective. Therefore there is a need to look for more general methods. The following result, due to Mittag-Leffler[3] tells us that, if we choose the 'singular portions' appropriately, we can always get a representation as above for all meromorphic functions on $\mathbb{C}$. The proof that we present here is simpler than the original one and is due to Weierstrass. The idea is to introduce 'convergence inducing summands' which are necessarily holomorphic functions so as not to disturb the pole set nor the singular parts.

Theorem 8.4.1 Mittag-Leffler (I-Version). *Let $P = \{b_1, b_2, \ldots\}$ be a (countable) discrete subset of $\mathbb{C}$. Let P_j be non zero polynomials without constant terms. Then there exists a meromorphic function f in $\mathbb{C}$ with pole set $P(f) = P$ and with the singular parts $P_j\left(\frac{1}{z-b_j}\right)$. Moreover, if h is any meromorphic function with $P(h) = P$ and with the corresponding singular parts $P_j\left(\frac{1}{z-b_j}\right)$, then h can be written in the form*

$$h(z) = \sum_j \left[P_j\left(\frac{1}{z-b_j}\right) - p_j(z)\right] + g(z), \tag{8.10}$$

where, g is an entire function and p_j are suitably chosen polynomial functions.

Proof: In this problem, if P is a finite set there is nothing to discuss. Also, we can, at will, add (or delete) a finite number of points to (or from) P without changing the nature of the problem. In particular, without loss of generality we may and will assume that $0 \notin P$ and P is infinite. Consider the Taylor expansion of $P_j\left(\frac{1}{z-b_j}\right)$ around the origin and let p_j be the partial sum of this expansion say, up to degree n_j. The idea is to choose n_j sufficiently large to suit our purpose. Consider the remainder term $f_j(z) = P_j\left(\frac{1}{z-b_j}\right) - p_j(z)$. We know that the Taylor's series for $P_j\left(\frac{1}{z-b_j}\right)$ is uniformly convergent on $|z| \leq |b_j|/2$. Hence, we can choose n_j so that

$$|f_j(z)| < 2^{-j}, \ \forall\ |z| \leq |b_j|/2, \ \forall\ j. \tag{8.11}$$

Now given any compact set K (since P is discrete), there exists a natural number $m = m(K)$ such that $K \subseteq \{z : |z| \leq |b_m|/2\}$. Hence, we can find some constants c_K such that the series $c_K + \sum_{j \geq m(K)} 2^{-j}$ serves as a majorant for the series $\sum_j (f_j/K)$. Thus

[3]Magnus G. Mittag-Leffler(1846-1927) was a Swedish mathematician, a most colorful personality, loved and respected by all. He was greatly influenced by Weierstrass in his approach. His main contribution is in the theory of functions. He played a great part in inspiring later research.

conditions (MF1) and (MF2) of normal convergence have been verified. It follows that the series $\sum_j f_j$ converges normally to a meromorphic function h with $P(h) = P$ with singular part at b_j equal to $P_j\left(\frac{1}{z-b_j}\right)$. This completes the first part of the theorem. For the second part, we have only to observe that the function $g(z) = f(z) - h(z)$ is an entire function. ♠

Example 8.4.1 Consider the case wherein a set $\{b_j\}$ has been given with the property that there exists an integer $k \geq 0$ such that $\sum_j |b_j|^{-k} = \infty$, and $\sum_j |b_j|^{-k-1} < \infty$. We want to find a meromorphic function f with its pole set equal to $\{b_j\}$ and singular parts $\frac{1}{z-b_j}$. So we take $P_j(z) = z$ for all j. We then have,

$$\frac{1}{z-b_j} = -\frac{1}{b_j}\left(1 + \frac{z}{b_j} + \cdots + \frac{z^n}{b_j^n} + \cdots\right).$$

Therefore we take, $n_j = k-1$ for all j, in the proof of the above theorem. Then

$$p_j(z) = \begin{cases} -\left(\frac{1}{b_j} + \frac{z}{b_j^2} + \cdots + \frac{z^{k-1}}{b_j^k}\right) & \text{if} \quad k > 0 \\ 0 & \text{if} \quad k = 0. \end{cases}$$

Hence as in the proof of the above theorem, we have,

$$f_j(z) = -\left(\frac{z^k}{b_j^{k+1}} + \frac{z^{k+1}}{b_j^{k+2}} + \cdots +\right).$$

Now, using Weierstrass' double series theorem, it can be directly verified that $\sum f_j(z)$ is normally convergent.

Work out the same problem taking $P_j(z) = z^2$, for all j. (It turns out that we can take $n_j = k-2$, if $k \geq 2$.)

Remark 8.4.1 What happens when consider the same problem on an arbitrary domain Ω in $\mathbb{C}$ instead of the whole plane? The above proof will run into difficulties in several places. We shall handle this problem in section 8.6, in an entirely different way.

Exercise 8.4

1. Find a meromorphic function f with $P(f) = \mathbb{Z}$ and principal parts $P_j(z)$ given by (i) z; (ii) jz for all $j \in \mathbb{Z}$.

2. As in the example (8.4.1) above, if $P = \{j^p \ : \ j \in \mathbb{N}\}$, where $p > 0$ is a fixed real number, determine $k = k_j$. Do the same for $P = \{e^j \ : \ j \in \mathbb{N}\}$.

8.5 Infinite Products

The study of poles was in a way simpler than the study of zeros. The sum $f_1 + f_2$ of two functions usually had the set of poles as the union of the set of poles of f_1 and f_2. On the other hand, we observe that the zero-set of the product $f_1 f_2$ is the union of the zero-sets of f_1 and f_2. When we wanted a meromorphic function with an infinite (discrete) set of poles we were naturally led to consider infinite sums and thereby to the theory of convergence of series. The analogous problem for the zero-sets leads us to the theory of infinite products.

Our aim here is to develop the theory of infinite products after Weierstrass, only to the extent that we need to discuss the product representation of meromorphic functions. The reader may refer to Hille's book (Vol I) or to any analytic number theory book, for more details.

Consider a sequence of complex numbers $\{c_k\}$ and let $C_n = \prod_{k=1}^n (1 + c_k)$. We would like to say that the infinite product

$$C = \prod_1^\infty (1 + c_k) \tag{8.12}$$

is convergent if the sequence $C_n = \prod_1^n (1 + c_k)$ is convergent. But wait a minute. Even if for some k we have, $c_k = -1$, then we would get the sequence to be convergent to zero. This will definitely make the definition too insensitive. Perhaps we can restrict ourselves to sequences where $c_k \neq -1$. This seems to be perfectly alright if our aim is to discuss sequences of points in $\mathbb{C}$ only. However, soon, we would like to study the product of sequences of functions also. It could be the case that even though our functions are not identically zero they may possess a number of zeros. Therefore, there is a need to select the definition carefully. As in the case of convergence of meromorphic functions, this difficulty is overcome by taking a middle course of allowing only finitely many zeros in any given sequence, and then 'ignore' them till we take the product and the limit. These ideas are made precise in the following:

Definition 8.5.1 For $1 \leq m < n$, let

$$C_{m,n} = \prod_m^n (1 + c_k).$$

We say the *infinite product(8.12) is convergent* if for some large m, $\lim_{n \longrightarrow \infty} C_{m,n}$ is a non zero complex number. Observe that a necessary condition that this limit exists is that $c_k \neq -1$ for large k. Of course, you can easily formulate and prove the Cauchy-type criterion for this convergence. By taking logarithms, it follows that the infinite product converges iff the sum

$$\sum_{m}^{\infty} \ln(1+c_k) \tag{8.13}$$

converges for large m. The following result is easy to prove.

Lemma 8.5.1 Let c_k be a sequence of non negative real numbers. Then (8.12) is convergent iff $\sum c_k$ is.

Proof: Put $a_n := 1 + c_1 + \cdots + c_n$. Then both the sequences $\{a_n\}$ and $\{C_{1,n}\}$ are monotonically increasing and we have

$$a_n \leq C_{1,n} \leq e^{a_n} \quad \forall\, n.$$

Hence by the so called Sandwich theorem, the lemma follows. ♠

Definition 8.5.2 We say that the *infinite product (8.12) is absolutely convergent* iff the infinite product

$$\prod_k (1+|c_k|)$$

is convergent .

The following two lemmas can be proved easily.

Lemma 8.5.2 Define $\Theta_{m,n} = \prod_{k=m}^{n}(1+|c_k|),\ \forall\, m \leq n$. Then

$$|C_{m,n+p} - C_{m,n}| \leq \Theta_{m,n+p} - \Theta_{m,n},\ \forall\, m < n,\ p \geq 1.$$

Lemma 8.5.3 An absolutely convergent product is convergent.

Definition 8.5.3 Let f_k be a sequence of continuous functions in a domain Ω and A be any subset of Ω. Put

$$\Phi_{m,n}(z) = \prod_{k=m}^{n}(1+f_k(z)).$$

We say that the *infinite product*

$$\prod_{k=1}^{\infty}(1+f_k) \tag{8.14}$$

is uniformly convergent on A iff

(0) for every $z \in A$, $\prod_1^\infty (1+f_k(z))$ is convergent;

(1) there exists $m(A)$ such that $f_n(z) \neq -1, \ \forall z \in A, \ n \geq m(A)$ and

(2) for every $\epsilon > 0$, there exists $N(\epsilon)$ such that $\forall \ n > N(\epsilon)$ and $m > m(A)$, we have,

$$|\Phi_{m,n}(z)|\,|\Phi_{n+1,n+p}(z) - 1| < \epsilon, \quad p \geq 1, \quad \forall \ z \in A.$$

If the product is uniformly convergent on each compact subset of Ω then we say that the *product is compactly convergent.*

Observe that, condition (1) for a compact subset $A = K$ implies that the bounded continuous functions $\Phi_{1,n}(z)$ may possibly vanish for some points on K but no product $\Phi_{m,n}(z)$ for $m > m(K)$, vanishes. Also, from (2) it follows that the sequences $\Phi_{m,n}(z)$ converge uniformly on K. Hence the sequence $\Phi_{1,n}(z)$ also converges uniformly on K. The following theorem is now a routine exercise. We suggest that the reader should write down complete details of the proof of it, before proceeding with further reading.

Theorem 8.5.1 *Let f_n be a sequence of holomorphic functions in Ω such that the product (8.14) is compactly convergent. Then the infinite product $f(z) = \prod_{n=1}^\infty (1+f_n(z))$ defines a holomorphic function in Ω. The infinite product is compactly convergent if the sum $\sum |f_n(z)|$ is. Moreover, the set of zeros of f is precisely equal to the union of the set of zeros of all f_n's, and the multiplicity at any of these zeros is precisely the sum of the multiplicities of f_n's.*

The analogue of Leibniz's rule applies to the derivative of an infinite product also, as seen below. The latter part of the theorem below tells you how take the logarithmic derivative of an infinite product.

Theorem 8.5.2 *Let $f(z) = \prod_{n=1}^\infty (1+f_n(z))$ be a convergent product of holomorphic functions. Define $g_n(z) = \sum_{k=1}^n \left[f_k'(z) \prod_{1\leq j\leq n,\ j\neq k} (1+f_j(z)) \right]$. Then $\lim_{n\longrightarrow\infty} g_n(z) = f'(z)$. Moreover, on any compact subset K on which $f(z) \neq 0$, we have,*

$$\frac{f'(z)}{f(z)} = \sum_n \frac{f_n'(z)}{1+f_n(z)}. \tag{8.15}$$

Proof: The first part follows directly theorem 8.1.3, since, g_n is the derivative of the sequence $\Phi_{1,n}$ which is uniformly convergent to f. If on a compact set K, f does not vanish then so do all the $\Phi_{n,1}$ and it follows that $\dfrac{g_n}{\Phi_{1,n}}$ converges uniformly to f'/f. But

$$\frac{g_n(z)}{\Phi_{1,n}(z)} = \sum_{j=1}^{n} \frac{f_j'(z)}{1 + f_j(z)}.$$

Hence the second result follows. ♠

With these preliminaries, let us now turn our attention to our goal. We start with an entire function f which does not vanish at all. One way to produce such a function is to take an entire function g and exponentiate:

$$f(z) = \exp\, g(z). \tag{8.16}$$

Indeed as seen in Exercise 4.11.1, this is the only way. For future use let us record this as a theorem.

Theorem 8.5.3 *An entire function $f(z)$ can be expressed as $f(z) = exp(g(z))$ for some other entire function iff it has no zeros.*

The next step is to consider entire functions which may have some zeros. Suppose we have an entire function f with the number of zeros finite, say, $a_0 = 0, a_1, a_2, \ldots, a_n$, with respective multiplicities $\mu_i \geq 0$. It follows that

$$f(z) = z^{\mu} \prod_{k=1}^{n} \left(1 - \frac{z}{a_k}\right)^{\mu_k} exp\,(g(z)) \tag{8.17}$$

When we have to deal with infinitely many zeros, you see that we need to have the notion of convergence of infinite products. Thus in (8.17) if we merely replace n by ∞, then we get the corresponding statement for infinite product provided the rhs makes sense. The first condition that is necessary is that the set of zeros should be discrete. Next, compact convergence of the infinite product should be ensured. By lemmas 8.5.1 and 8.5.3, this is so if we can ensure that $\sum_n(|\mu_n/a_n|)$ is convergent. Since this is too stringent a condition, the answer so far is not quite satisfactory. So, as in the case of Mittag-Leffler's theorem, we should try to modify the product by introducing the convergence inducing factors. Such a modification was found by Weierstrass.

Let us introduce the notation:

$$\begin{aligned} E(z,0) &= 1 - z & &\text{and} \\ E(z,m) &= (1 - z)exp[p_m(z)], & &m \geq 1, \end{aligned}$$

where for $m \geq 1,\ p_m(z)$ is the m^{th} partial sum of the Maclaurin series for

$$ln\left(\frac{1}{1-z}\right) = \sum_1^\infty \frac{z^n}{n}.$$

Lemma 8.5.4 For $|z| \leq 1$, we have, $|E(z,m) - 1| \leq |z|^{m+1}$ for all $m \geq 0$.

Proof: For $m = 0$, is is obvious. So, let $m \geq 1$. Observe first that $E'(z,m) = -z^m exp[p_m(z)]$. Let is write $E(z,m) := \sum b_n z^n$. Then that $\sum_{n=1}^\infty n b_n z^{n-1} = -z^m exp[p_m(z)]$. Upon comparing the coefficients on either side, we get, $b_1 = \cdots = b_m = 0$ and $b_n < 0$ for all $n > m$. Also, clearly $b_0 = 1$. Finally, $0 = E(z,1) = 1 + \sum_{n>m} b_n$ and hence, $\sum_{n>m} b_n = -1$, and hence $\sum_{n>m} |b_n| = 1$. Therefore, for all $|z| = 1$, we have, $|\sum_{n>m} b_n z^{n-m-1}| \leq 1$. By maximum principle, the same holds for all $|z| \leq 1$. Therefore, for $|z| \leq 1$,

$$|E(z,m) - 1| \leq \sum_{n>m} |b_n||z|^m = |z|^{m+1} \sum_{n>m} |b_n||z|^{n-m-1} \leq |z|^{m+1}$$

as claimed. ♠

Theorem 8.5.4 Weierstrass' Theorem (I-Version): *Let $a_0 = 0, a_1, a_2, \ldots, a_n, \ldots$ be a sequence of distinct points in $\mathbb{C}$ without any limit points and let $\{\mu_j\}$ be a sequence of non negative integers. Then there exists an entire function having a zero of multiplicity μ_j at each a_j and no other zeros. Moreover, any other entire function h with the same property will be of the form*

$$h(z) = exp(g(z)) z^{\mu_0} \prod_{n=1}^\infty \left(E\left(\frac{z}{a_n}, m_n\right)\right)^{\mu_n}$$

for some entire function $g(z)$ and some suitably chosen sequence of integers m_n such that $\sum \mu_n \left|\frac{z}{a_n}\right|^{m_n+1}$ is uniformly convergent in every closed disc $|z| \leq R$.

Proof: By the above lemma, we have

$$\left|E\left(\frac{z}{a_n}, m\right) - 1\right| \leq \left|\frac{z}{a_n}^{m+1}\right|, \quad |z/a_n| \leq 1.$$

Without loss of generality, we may and will assume that $|a_n| \leq |a_{n+1}|$ for all n. Hence given $R > 0$, we can find N such that for all $n > N$, we have $|a_n| > R$, so that the above inequality is valid for all $n > N$. Now, in order to prove the absolute and uniform convergence of the product in $|z| \leq R$, it suffices to find a sequence of integers m_n such that $\sum \mu_n \left|\frac{z}{a_n}\right|^{m_n+1}$ converges uniformly on $|z| \leq R$ for $R > 0$. We claim that it suffices to choose m_n so that

$$m_n + 1 \geq \ln(n^2 \mu_n), \quad \forall \;\; n.$$

For, then given $R > 0$, first choose $N(R)$ so that $|a_n| > eR$, for all $n > N(R)$ so that,

$$\mu_n \left|\frac{R}{a_n}\right|^{m_n+1} < \frac{\mu_n}{e^{m_n+1}} < \frac{\mu_n}{n^2 \mu_n} = \frac{1}{n^2}.$$

Thus, $\sum_n \mu_n \left|\frac{R}{a_n}^{m_n+1}\right|$ is convergent and serves as a majorant for $\sum_n \mu_n (E(\frac{z_n}{a_n}, m_n) - 1)$. As indicated already, this proves the absolute and uniform convergence of the product. For the last part of the theorem, we simply appeal to theorems 8.5.1 and 8.5.3. ♠

The following corollary is immediate.

Corollary 8.5.1 Every meromorphic function is a quotient of two entire functions. [Thus $\mathcal{M}(\mathbb{C})$ is the quotient field of $\mathcal{H}(\mathbb{C})$.]

Proof: Let f be a meromorphic function. It is enough to show that there exists an entire function g such that gf is also entire. For this we choose g to be an entire function prescribed by the above theorem where the set of zeros of g coincides with the set of poles of f with the multiplicities of these zeros for g being the order of the respective poles for f. ♠

Example 8.5.1 Product Representation of $\sin \pi z$: It follows immediately that

$$\sin \pi z = e^{g(z)} z \prod_{n \neq 0} \left[\left(1 - \frac{z}{n}\right) e^{z/n}\right]. \tag{8.18}$$

for some entire function $g(z)$. To determine $g(z)$, we differentiate the logarithm of the above expression.

$$\pi \cot \pi z = g'(z) + \frac{1}{z} + \sum_{n \neq 0} \left(\frac{1}{z-n} + \frac{1}{n}\right) = g'(z) + \frac{1}{z} + \sum_{n \geq 1} \left(\frac{2z}{z^2 - n^2}\right).$$

From (8.7), it follows that $g'(z) = 0$. Hence, $g(z)$ is a constant. The value of this constant is easily determined to be equal to π by taking the limit of $\frac{\sin \pi z}{z}$ as $z \longrightarrow 0$.

Thus we have,

$$\boxed{\sin \pi z = \pi z \prod_{n \neq 0} \left[\left(1 - \frac{z}{n}\right) e^{z/n} \right]} \tag{8.19}$$

In the next section we shall see some far reaching generalization of Weiestrass's as well as Mitteg-Leffler's theorems by a different approach.

Exercise 8.5

1. For what values of $p > 0$, does the product $\prod_1^\infty (1 + n^{-p})$ converge?

2. Show that $\prod_1^\infty \left(1 + \frac{\imath}{n}\right)$ diverges whereas $\prod_1^\infty |1 + \frac{\imath}{n}|$ converges. (Observe the definition of the absolute convergence of the product once again.)

3. Discuss the convergence of $\prod_1^\infty \left(1 + \frac{(-1)^{n-1}}{n}\right)$.

8.6 Runge's Approximation Theorem

Let us begin with a

Question: Let K be a compact subset of $\mathbb{C}$. Which continuous functions $f : K \longrightarrow \mathbb{C}$ are uniform limits on K of polynomials?

Example 8.6.1 As an amusing illustration of application of maximum modulus principle, let us prove the following: *There is no sequence of polynomials $p_n(z)$ which uniformly approximates the function $1/z$ on the unit circle.*

Assuming on the contrary, let $\{p_n\}$ be a sequence of polynomials which uniformly converges to $1/z$ on $\mathbb{S}^1$. In particular it follows that the sequence is uniformly Cauchy on $\mathbb{S}^1$. By maximum principle this implies that the sequence is uniformly Cauchy on the unit disc itself. Therefore, by Weierstrass convergence theorem 8.1.3, the sequence converges to a holomorphic function f on the unit disc. Since $f(z) = 1/z$, on $\mathbb{S}^1$, this leads to the absurd conclusion that 0 is a removable singularity of $1/z$.

More generally, it follows that in order to seek an affirmative answer to the above question, two conditions on such a function f are seen to be necessary:

(1) By Weierstrass's theorem, 8.1.3, being a uniform limit of polynomial functions f has to be holomorphic in $int\, K$.
(2) Suppose U is a bounded component of $\mathbb{C}\setminus K$. (This will exist if $\mathbb{C}\setminus K$ is disconnected, since K is compact.) Clearly $\partial U \subset K$. If a sequence P_n of polynomials is uniformly convergent on K to f, then by the maximum principle, it follows that $\{P_n(z)\}$ is a uniformly Cauchy on U and hence will converge uniformly in U to a limit function which is then holomorphic on U and of course agrees with f on K. Thus f should be holomorphic on all such components U of $\mathbb{C} \setminus K$.

The first condition is within our control as soon as the function f is given on K. But the second condition says that we must be able to extend f holomorphically on all bounded components of $\mathbb{C} \setminus K$, which information, apriori, may not be available to us. Therefore, it seems natural to avoid this case(2) by imposing a connectivity condition on the complement of K and here is such an answer.

Theorem 8.6.1 (Mergelyan-1950) *Let K be a compact subset of $\mathbb{C}$ such that $\mathbb{C} \setminus K$ is connected. Let $f : K \longrightarrow \mathbb{C}$ be a is continuous function on K and holomorphic in $int\, K$. Then f is the uniform limit of a sequence of polynomials on K.*

Remark 8.6.1 We shall not prove this theorem here. See [Ru2] for example. We shall prove something similar to the above, restricted in one direction but more general in another. Namely, we shall restrict ourselves to the case when f is holomorphic on the whole of K. On the other hand, we shall not impose the connectivity condition and work with rational functions rather than polynomials. Later we shall specialize to see what happens when connectivity condition is imposed.

Let $\mathcal{H}(K)$ denote the space of all holomorphic functions on K, i.e., $f \in \mathcal{H}(K)$ if it is the restriction of a holomorphic function in an open set containing K. In what follows, we take the supremum norm $\| - \|_K$ and the corresponding topology on $\mathcal{H}(K)$.

Theorem 8.6.2 Runge's Theorem I-Version *Let K be a compact subset of $\mathbb{C}$ and $\{U_j\}$ be the set of connected components of $\hat{\mathbb{C}} \setminus K$. Let $B \subset \hat{\mathbb{C}}$ be such that $B \cap U_j \neq \emptyset$ for all j. Then any $f \in \mathcal{H}(K)$ can be approximated uniformly on K by rational functions whose pole set is contained in B.*

Corollary 8.6.1 Let K be a compact subset of $\mathbb{C}$ such that $\mathbb{C} \setminus K$ is connected. Then any $f \in \mathcal{H}(K)$ can be uniformly approximated on K by polynomial functions.

Proof: It follows that $\hat{\mathbb{C}} \setminus K$ is also connected. So, we can choose $B = \{\infty\}$. Clearly a rational function with ∞ as the only pole is a polynomial function. ♠

Remark 8.6.2 Thus the above corollary is only a bit weaker than Theorem 8.6.1. In some sense, this result says that there are **not** too many functions in $\mathcal{H}(K)$. On the other hand, this is one result which has been put to use in constructing bizarre types of functions on $\mathbb{C}$. Here is an example of this.

Example 8.6.2 Define

$$A_n = \{x + \imath y \ : \ |x| \leq n, \ 0 \leq y \leq n\}, \quad B_n = \cup\{x + \imath y \ : \ |x| \leq n, \ -n \leq y \leq -1/n\}.$$

For each n, take $K = A_n \cup B_n$ and $\Omega = \mathbb{C}$ in corollary 8.6.1, to find a holomorphic function f_n which is an approximation to the function ϕ_n such that $\phi_n(z) = 1, z \in A_n$ and $\phi_n(z) = 0, z \in B_n$. If f is the point-wise limit of f_n then $f(z) = 1$ for all z such that $\Re(z) \geq 0$ and $f(z) = 0$ for $\Re(z) < 0$. Observe that the convergence of $f_n \to f$ is not uniform on compact sets.

The proof of theorem 8.6.2 can be broken up into two major steps:

Lemma 8.6.1 Any $f \in \mathcal{H}(K)$ can be approximated by a linear combination of $\dfrac{1}{z-\alpha}, \alpha \notin K$.

Lemma 8.6.2 Pole Shifting Let $K \subset U \subset \hat{\mathbb{C}}$, where U is open and K is compact. Suppose p, q are in the same component V of $\hat{\mathbb{C}} \setminus K$ and $p \neq \infty$. Then any polynomial in $\dfrac{1}{z-p}$ can be approximated on K by rational functions having pole at q alone.

The theorem follows immediately from these two lemmas.

Lemma 8.6.3 Let $g : K \times [a, b] \longrightarrow \mathbb{C}$ be a continuous function where K is a compact set. Given $\epsilon > 0$, there exists a partition of the interval $[a, b]$ say $a = t_1 < t_2 < \cdots < t_{n+1} = b$ such that for every $j = 1, 2, \ldots, n-1$,

$$|g(z, t) - g(z, t_j)| < \epsilon, \quad \forall \ \ z \in K \ \ \& \ \ t_j \leq t \leq t_{j+1}. \tag{8.20}$$

Proof: This follows easily by standard arguments using the uniform continuity. (See for example Lebesgue Covering Lemma 7.3.2.)

Proof lemma 8.6.1: Let $K \subset U \subset \mathbb{C}$, where U is open and $f \in \mathcal{H}(U)$. By lemma 7.4.1, there exists a cycle ω in $U \setminus K$ such that for every $z \in K$,

$$\eta(\omega, z) = 1 \ \& \ f(z) = \frac{1}{2\pi\imath}\int_{\omega}\frac{f(\xi)}{\xi - z}d\xi. \tag{8.21}$$

Recall that a cycle is a finite sum of contours and each contour is made up of finitely many piecewise smooth curves. Therefore, the integral in (8.21) is a finite sum of such integrals over smooth curves. So, it suffices to assume that ω is itself is a smooth curve defined on an interval $[a, b]$ and approximate the integral

$$\int_{\omega}\frac{f(\xi)}{\xi - z}d\xi, \ \ z \in K.$$

Take

$$g(z,t) = \frac{f(\omega(t))}{\omega(t) - z}$$

and find a partition of the interval as above so that 8.20 holds. Put

$$P(z) = \frac{1}{2\pi\imath}\sum_{j=1}^{n}\frac{f(\omega(t_j))}{\omega(t_j) - z}(\omega(t_{j+1}) - \omega(t_j)).$$

Then P has its pole set in $\{\omega(t_j)\}_{1\le j\le n} \subset U \setminus K$. Also,

$$\begin{aligned}
|f(z) - P(z)| &= \left|\frac{1}{2\pi\imath}\int_{\omega}\frac{f(\xi)}{\xi - z}\,d\xi - P(z)\right| \\
&= \frac{1}{2\pi}\left|\sum_{j=1}^{n}\int_{t_j}^{t_{j+1}}\left(\frac{f(\omega(t))}{\omega(t) - z} - \frac{f(\omega(t_j))}{\omega(t_j) - z}\right)\omega'(t)dt\right| \\
&< \frac{\epsilon\, l(\omega)}{2\pi}
\end{aligned}$$

where $l(\omega)$ denote the length of the curve ω. This complete the proof of lemma 8.6.1.♠

Lemma 8.6.4 Suppose $d(p, q) < d(K, p)$. Then $\dfrac{1}{z - p}$ can be approximated on K by a polynomial in $\dfrac{1}{z - q}$.

Proof: We have, for $z \in K$,

$$\frac{|p - q|}{|z - q|} \le \frac{|p - q|}{d(K, q)} < 1.$$

Therefore, the geometric series in $\dfrac{|p - q|}{|z - q|}$ converges and we have,

$$\begin{aligned}\frac{1}{z-p} = \frac{1}{z-q-(p-q)} &= \frac{1}{z-q}\left(\frac{1}{1-\frac{p-q}{z-q}}\right) \\ &= \frac{1}{z-q}\left(1+\frac{p-q}{z-q}+\cdots+\left(\frac{p-q}{z-q}\right)^j+\cdots\right).\end{aligned}$$

Proof of lemma 8.6.2: Consider the first case when V is a component not containing ∞. Choose a path τ from p to q in V. Let $\delta > 0$ be the distance of τ from K. Choose (distinct) points $p = p_1, \ldots, p_n = q$ on τ such that $|p_{k+1} - p_k| < \delta$. Then for all $z \in K$, $|p_{k+1}-p_k| < d(K, p_k)$ for all k. By the above lemma 8.6.4, $\dfrac{1}{z-p_1}$ can be approximated by a polynomial in $\dfrac{1}{z-p_2}$. Therefore, by taking powers and their finite linear combinations, this means that any polynomial in $\dfrac{1}{z-p_1}$ can also be approximated by a polynomial in $\dfrac{1}{z-p_2}$. Now a simple induction completes this case.

Next, suppose V contains ∞. If $q \neq \infty$ then we can choose a path in V from p to q and not passing through ∞. (It is crucial here that K is compact.) We can then proceed exactly as in the preceding paragraph. Finally, consider the case $q = \infty$. Then there exists $q' \in V$, $q' \neq \infty$ such that $|q'| > 2|z|$ for all $z \in K$. By the previous case, the pole can be shifted to q' first. But then we have, $|z/q'| < 1/2$ and hence

$$\frac{1}{z-q} = -\frac{1}{q}\left(1+\frac{z}{q}+\cdots+\frac{z^j}{q^j}+\cdots\right).$$

This shows that all powers of $\dfrac{1}{z-q}$ can now be approximated by polynomials. This completes the proof lemma 8.6.2 and thereby that of theorem 8.6.2. ♠

Remark 8.6.3 If we could choose the pole set B to be disjoint from an open set Ω, we can immediately say that $f \in \mathcal{H}(K)$ can be approximated by holomorphic functions in Ω. The following result now gives a purely topological criterion which ensures this.

Theorem 8.6.3 *Let $\Omega \subset \mathbb{C}$ be an open set and $K \subset \Omega$ be compact. Then the following three conditions are equivalent:*
(1) *Every $f \in \mathcal{H}(K)$ can be approximated by holomorphic functions on Ω.*
(2) *$\Omega \setminus K$ has no connected component whose closure in Ω is compact.*
(3) *Every bounded component of $\mathbb{C} \setminus K$ meets $\mathbb{C} \setminus \Omega$.*

Proof: (1) $\Longrightarrow$ (2) Suppose U is a component of $\Omega \setminus K$ whose closure in Ω is compact. For some $b \in U$, consider $f(z) = \dfrac{1}{z-b}$ which is holomorphic on K. Let f_n be a sequence of holomorphic functions on Ω converging uniformly on K to f. Given ϵ there exists n such that $|f_n - f_{n+k}|_K < \epsilon$ for all $k \geq 1$. Since the boundary of U is contained in K, by maximum principle, it follows that $|f_n - f_{n+k}|_L < \epsilon$ for all compact subsets L of U. This means $\{f_n\}$ is uniformly Cauchy on compact subsets of U. Therefore, $\{f_n\}$ is compactly convergent on U to a function g which is, by Weierstrass theorem, holomorphic on U. But $g = f$ on $\partial U \subset K$. Since ∂U cannot be a discrete set, this means $g = f$ on U which means b is a removable singularity of f. This is a contradiction.

(2) $\Longrightarrow$ (3) : Assume (3) is not true. Let U be a bounded component of $\mathbb{C} \setminus K$ which does not meet $\mathbb{C} \setminus \Omega$. This is the same as saying $U \subseteq \Omega$. Observe that the closure $\overline{U}$ of U in $\mathbb{C}$ is compact and $\partial U := \overline{U} \setminus U \subset K$. Therefore, the closure of U in Ω is equal to $\overline{U}$ and hence compact. This contradicts (2). This completes the proof.

(3) $\Longrightarrow$ (1) Choose the set B such that $B \cap \Omega = \emptyset$. Apply theorem 8.6.2. ♠

We shall now take up the task of moving from a compact set K to any open set.

Theorem 8.6.4 *Let $\Omega \subset \mathbb{C}$ be open. Then there exist a sequence of compact sets K_n in Ω such that*

(a) $K_n \subset int\, K_{n+1}$ *for all* n;

(b) $\cup_n K_n = \Omega$ *and*

(c) *each component of* $\mathbb{C} \setminus K_n$ *meets* $\mathbb{C} \setminus \Omega$.

Proof: Let $B_n(0)$ denote the closed ball of radius n with centre 0. Define

$$K_n = B_n(0) \cap \{z \ : \ d(z, \mathbb{C} \setminus \Omega) \geq 1/n\}.$$

Then clearly, being a closed subset of $B_n(0)$, each K_n is compact. Also, it is fairly easy to see that $K_n \subset int\, K_{n+1}$ and $\cup_n K_n = \Omega$. Now fix n and suppose U is a bounded component of $\mathbb{C} \setminus K_n$. Observe that $\{z \ : \ |z| > n\}$ is connected and hence is contained in the unbounded component of $\mathbb{C} \setminus K_n$. It follows that U is contained in the interior of $B_n(0)$. Hence, U must contain a point z such that $d(z, \mathbb{C} \setminus \Omega) < 1/n$. This means that there exists $w \in \mathbb{C} \setminus \Omega$ such that $|z - w| < 1/n$. But then for all points y in the line segment $[z, w]$, we have, $d(y, \mathbb{C} \setminus \Omega) < 1/n$. Hence $[z, w] \cap K_n = \emptyset$. Since U is a component of $\mathbb{C} \setminus K_n$, it follows that $[z, w] \subset U$. In particular, $w \in U$ and hence $U \cap \mathbb{C} \setminus \Omega \neq \emptyset$. Hence condition (c) is verified.

Theorem 8.6.5 Runge's Theorem (II Version): *Let Ω be an open set in $\mathbb{C}$ and A be a subset of $\hat{\mathbb{C}}$ which meets each component of $\hat{\mathbb{C}} \setminus \Omega$. Then each $f \in \mathcal{H}(\Omega)$ can be approximated on Ω uniformly on compact sets by rational functions having their pole set contained in A. In the special case, when $\hat{\mathbb{C}} \setminus \Omega$ is connected, we can do the same with polynomial functions.*

Proof: Let $\{K_n\}$ be a sequence of compact sets as given in theorem 8.6.4. By condition (c) therein, it follows that each component C of $\hat{\mathbb{C}} \setminus K_n$ meets $\hat{\mathbb{C}} \setminus \Omega$ and hence contains a component of the latter. Therefore $A \cap C \neq \emptyset$. Hence, by the I-version of Runge's theorem 8.6.2, there exists a rational function R_n such that $|R_n - f|_{K_n} < \frac{1}{n}$. Let now L be any compact subset of Ω. There exists n_1 such that $L \subset K_{n_1}$. So, given $\epsilon > 0$, choose $n > \max\{n_1, 1/\epsilon\}$. Then

$$|R_k - f|_L \leq |R_k - f|_{K_k} < \epsilon, \ \forall \ k > n.$$

This means that the sequence $\{R_n\}$ compactly converges to f on Ω as required. The latter part follows as before, since we can take $A = \{\infty\}$. ♠

Remark 8.6.4 We shall now give a number of applications of this theorem. The power series representation of a holomorphic function is a very special way of approximating it by polynomials, albeit restricted only over certain discs. Runge's theorem can be viewed as a very sweeping generalization of this. For instance, we can now say that a holomorphic function in a simply connected region, can be uniformly approximated by polynomials. This is of course not as strong as power series representation. Similarly we have:

Laurent Series: Recall that if we have a holomorphic function f defined in an annular region Ω, then it has a Laurent series representation. This just means that f can be approximated by Laurent polynomials in $z-a$, where a is the centre of the annulus. This result can be generalized as follows. Let us say Ω is an annular like region, if $\Omega = U_1 \setminus \overline{U}_2$ where U_j are simply connected regions such that $\overline{U}_2 \subset U_1$. Then any $f \in \mathcal{H}(\Omega)$ can be approximated by Laurent polynomials in $z - a$ for any point $a \in U_2$. All that we do is to take $A = \{a, \infty\}$ in the theorem 8.6.5.

Any property preserved under uniform convergence and which holds for approximating functions will also hold for approximated functions. Therefore, we can anticipate quite a few applications of Runge's theorem. A typical illustration of this is:

Homology Form of Cauchy Theorem Recall the statement from theorem 7.4.1. The proof of '$(ii) \Longrightarrow (i)$' can be given as follows: Take $A = \hat{\mathbb{C}} \setminus \Omega$ in theorem 8.6.5, to get approximations to f by rational functions R_n with poles in A. Condition (ii) implies that $\int_\gamma R_n(z)dz = 0$. Upon taking limit as $n \longrightarrow \infty$, we get $\int_\gamma f(z)dz = 0$.

The gain here is not much since lemma 7.4.1, which is the main ingredient in the proof of theorem 7.4.1 that we have given in Ch. 7, is also used in the proof of Runge's theorem. Next, we shall fulfil our promise of giving an improved version of Mittag-Leffler theorem.

Theorem 8.6.6 General Form of Mittag-Leffler *Let B be a discrete subset of an open set Ω in $\mathbb{C}$. To each $b \in B$ let P_b be a non zero polynomial without a constant term. Then there exists a meromorphic function f on Ω with its pole set precisely B and its principal part at b equal to $P_b\left(\dfrac{1}{z-b}\right)$. Moreover, any other meromorphic function g with this property differs from f by a holomorphic function on Ω.*

Proof: Write $\Omega = \cup_n K_n$ where K_n are as in theorem 8.21. B being a discrete subset of an open set in a euclidean space, is countable. Hence, we can enumerate it in some way $B = \{b_j\}$. It follows that for each n there exists $m(n)$ such that $b_j \notin K_n$ for $j \geq m(n)$. Then $P_{b_j}\left(\dfrac{1}{z-b_j}\right)$ is holomorphic on K_n and hence there exist a holomorphic function g_j on Ω (in fact, a rational function with pole set in $\mathbb{C} \setminus \Omega$) such that

$$\left| P_{b_j}\left(\frac{1}{z-b_j}\right) - g_j(z) \right| < \frac{1}{2^j},$$

for all $j \geq m(n)$ and for all $z \in K_n$. Take $f_j(z) = P_{b_j}\left(\dfrac{1}{z-b_j}\right) - g_j(z)$. Then $\sum_j f_j(z)$ satisfies the conditions for normal convergence, and hence defines a meromorphic function f on Ω with its principle parts precisely $P_{b_j}\left(\dfrac{1}{z-b_j}\right)$.

The last statement of the theorem is, of course, obvious. ♠

Remark 8.6.5 If we closely follow the proof of Runge's theorem, rather than the final statement it becomes clear that it is further possible to choose g_j as polynomials. Compare the direct proof that we have given of Mittag-Lefter theorem.

Theorem 8.6.7 Weierstrass's theorem II-Version: *Given a discrete subset Z of an open set Ω, there exists a holomorphic function f on Ω having zeros precisely at Z with pre-assigned multiplicities.*

Proof: Given $Z = \{\zeta_j\} \subset \Omega$, and positive integers n_j, we would like to consider the function $f(z) = \prod_{j=1}^{\infty}(z-\zeta_j)^{n_j}$ but for the fact that the infinite product as such, may not converge. Hence we must multiply this by suitable convergence inducing factors. Thus we seek holomorphic functions $h_j(z)$ such that $\prod_j (z-\zeta_j)^{n_j}e^{h_j(z)}$ is convergent. Let us see how far this will work. We write $\Omega = \cup K_n$. Let m_r be the increasing sequence such that $\zeta_j \notin K_r$ for $j > m_r$. Now, if $\ln(z-\zeta_j)$ makes sense on K_r, then by 8.6.1, we can choose h_j such that $|\ln(z-\zeta_j) - h_j(z)| < \epsilon/2^j$. This is so, if ζ_j is in an unbounded component of $\mathbb{C}\setminus K_r$. We are in trouble otherwise. Say, now that ζ_j is in a bounded component U of K_r. Then U meets $\mathbb{C}\setminus\Omega$ and hence we can choose $\beta_j \in U \cap (\mathbb{C}\setminus\Omega)$. In this case, it follows that $\ln\left(\frac{z-\zeta_j}{z-\beta_j}\right)$ makes sense on K_r. Therefore, we should modify each such factor by a multiple of $(z-\beta_j)^{-n_j}$, so that the above argument works. Thus, the correct statement would be that to each j, we can find some $\beta_j \notin Z$ and an integer m_j (which is either zero or $= n_j$ and holomorphic functions h_j such that

$$\prod_j \left[(z-\zeta_j)^{n_j}(z-\beta_j)^{-m_j}e^{h_j(z)}\right]$$

is convergent to a holomorphic function on Ω with the prescribed zeros. ♠

Just as in the case when $\Omega = \mathbb{C}$, we also have:

Corollary 8.6.2 Every meromorphic function on a domain Ω is the quotient of two holomorphic functions. In other words, the quotient field of the ring $\mathcal{H}(\Omega)$ is the field of meromorphic functions $\mathcal{M}(\Omega)$.

Corollary 8.6.3 Every non empty open set is the natural domain of a non constant holomorphic function.

Proof: We first take note of a set topological fact: if Ω is an open set in $\mathbb{C}$, then there exists a discrete subset B in Ω such that its closure in $\overline{\Omega}$ contains $\partial\Omega$. (For example, if K_n are compact subsets of Ω as in theorem 8.6.4, take A_n to be all points $x + \imath y \in \Omega\setminus K_n$ where $2^n x$ and $2^n y$ are both integers, and put $B = \cup_n A_n$.) Now by Weierstrass's theorem, there exists $f \in \mathcal{H}(\Omega)$ with its zero set precisely equal to B. If f has an extension across Ω, i.e, if f is defined and holomorphic in a neighbourhood of some $\zeta \in \partial\Omega$, then since ζ is a limit point of B, it follows by the identity theorem that f is identically zero which is absurd. ♠

Finally we have the following strong form of Weierstrass's theorem which is completely similar to Mitteg-Leffler theorem.

Theorem 8.6.8 Weierstrass' theorem III-Version: *Let $\{b_j\}$ be a discrete subset of a domain Ω. Given arbitrary polynomials p_j, there exists a holomorphic function f on Ω such that for every j its power series representation around b_j coincides with $p_j(z - b_j)$ up to $deg\, p_j$ terms.*

Proof: First consider the case when each $p_j = \lambda_j$ is a constant. In this case, first, using Weierstrass's theorem, find a holomorphic function g such that b_j is a simple zero for each j. Then $g'(b_j) \neq 0$. Now using Mitteg-Leffler, find a meromorphic function h with a simple pole at each b_j and with singular parts, $\dfrac{\lambda_j}{g'(b_j)(z - b_j)}$. Now $f = gh$ is as required.

The general case is similar to this and resembles the way one writes the inverse of a formal power series with constant term non zero. Put $p_j(z) = \sum_{r=0}^{m_j} \alpha_{jr} z^r$. We first choose a holomorphic function g on Ω which has a zero of order $m_j + 1$ at b_j. We have to determine certain polynomials $q_j(z) = \sum_r \beta_{j,r} z^{m_j+1-r}$ such that if h is a meromorphic function with its singular part $q_j\left(\dfrac{1}{z - b_j}\right)$, then $f = gh$ will be the require function. Let $g(z) = \sum_{r>m_j} \lambda_{j,r}(z - b_j)^r$. We have to determine q_j by the requirement that

$$\left(\frac{\beta_{j,m_j+1}}{z - b_j} + \cdots + \frac{\beta_{j,1}}{(z - b_j)^{m_j+1}}\right)\left(\lambda_{j,m_j+1}(z - b_j)^{m_j+1} + \cdots\right)$$
$$= \alpha_{j,0} + \alpha_{j,1}(z - b_j) + \cdots + \alpha_{j,m_j}(z - b_j)^{m_j} + \cdots .$$

Using the single fact that $\lambda_{j,m_j+1} \neq 0$, this can be solved for $\beta_{j,1}, \beta_{j,2} \ldots, \beta_{j,m_j}$ successively. ♠

Exercise 8.6

1. Let $\Omega \subset \Omega' \subset \mathbb{C}$ be open subsets. Then the following two conditions are equivalent.
   (i) The restriction map $\rho : \mathcal{H}(\Omega') \longrightarrow \mathcal{H}(\Omega)$ has dense image.
   (ii) $\Omega' \setminus \Omega$ has no compact connected components.

2. Let $\Omega' \subset \mathbb{C}$ be an open set. For $f_1, \ldots, f_k \in \mathcal{H}(\Omega')$, let

   $$\Omega = \{z \in \Omega' \ : \ |f_j(z)| < 1, \quad 1 \leq j \leq k.\}$$

   Then the restriction map $\rho : \mathcal{H}(\Omega') \longrightarrow \mathcal{H}(\Omega)$ has dense image.

3. For any compact set $K \subset \Omega$, let $\widehat{K}$ be the union of K and all bounded components of $\mathbb{C} \setminus K$ contained in Ω. Show that
   (a) $\widehat{K}$ is compact.

(b) $z \in \widehat{K} \Longleftrightarrow$ for all $f \in \mathcal{H}(\Omega)$, we have, $|f(z)| \leq sup\{|f(\zeta)| \, : \, \zeta \in K\}$.
(c) If $K_1 \subset K_2$ are compact subsets of Ω then $\widehat{K_1} \subset \widehat{K_2}$.
(d) If $K_j, j = 1,2$ are compact subsets of Ω such that $\widehat{K}_j = K_j, j = 1,2$ then $\widehat{K_1 \cap K_2} = K_1 \cap K_2$. (For more on this important notion see the book of R. Narasimhan [N].)

4. Let K be any subset of $\mathbb{C}$ and $f : K \to \mathbb{C}$ be a continuous function. Call f locally extendable to a holomorphic function, if for each $z \in K$, there is $r > 0$ and a holomorphic function g on $B_r(z)$ such that $f = g$ on $B_r(z) \cap K$. Show that if K is compact then $\mathcal{H}(K)$ is equal to the set of all continuous functions on K which are locally extendable to a holomorphic function.

5. Let K be a compact subset of $\mathbb{C}$. Show that $\mathbb{C} \setminus K$ is connected iff for each $\alpha \in \mathbb{C} \setminus K$, $\ln(z - \alpha)$ has an (analytic) continuous branch over K. (Indeed, you may first observe that having an analytic branch of $\ln(z - \alpha)$ is equivalent to having a continuous branch of the same on K. See the previous exercise.)

6. Let $K_j, j = 1,2$, be compact subsets of $\mathbb{C}$ such that $\mathbb{C} \setminus K_j$ are connected. Suppose further that $K_1 \cap K_2 \neq \emptyset$ is connected. Then $\mathbb{C} \setminus K_1 \cup K_2$ is connected. (Hint: Use the previous exercise.)

8.7 The Gamma Function

The simplest entire function with the set of zeros as integers is $\sin \pi z$. Due to the functional properties such as periodicity etc. that this function enjoys, and its importance in Mathematics as a whole, it is considered as an elementary function. Closely related is another entire function which has simple zeros at all negative integers:

$$G(z) = \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) \exp(-z/n) \tag{8.22}$$

Observe that $G(-z)$ is then an entire function with its zero set as the set of positive integers and is as simple as G itself.

Now, using the product representation 8.19 for $\sin \pi z$, we easily see that

$$zG(z)G(-z) = \sin \pi z. \tag{8.23}$$

In order to study the functional properties of $G(z)$, we consider the function $G(z-1)$, which has its zero set as the set of non positive integers. It follows that

$$G(z-1) = \exp(\gamma(z))zG(z) \tag{8.24}$$

for some entire function $\gamma(z)$, which is to be determined. Take the logarithmic derivative of (8.24) to obtain

$$\sum_{n=1}^{\infty}\left(\frac{1}{z-1+n} - \frac{1}{n}\right) = \gamma'(z) + \frac{1}{z} + \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right) \tag{8.25}$$

Replacing n by $n+1$ on the lhs and observing that $\sum_{n=1}^{\infty}\left(\frac{1}{n+1} - \frac{1}{n}\right) = -1$, it follows that $\gamma'(z) = 0$. Therefore, $\gamma(z) = \gamma$ is a constant. To find the value of this constant, we observe that $G(0) = 1$ and put $z = 1$ in 8.24. It follows that

$$\gamma = -\ln G(1) = \lim_{n\longrightarrow\infty}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln n\right). \tag{8.26}$$

This real number is the celebrated *Euler's constant.* Its value has been calculated to several decimal places ($\sim .57722\cdots$). *It is not known whether γ is rational or irrational, algebraic or transcendental!*

Definition 8.7.1 The gamma function is defined by the formula:

$$\Gamma(z) := \exp(-\gamma z)z^{-1}G(z)^{-1} = \frac{\exp(-\gamma z)}{z}\prod_{n=1}^{\infty}\frac{\exp(z/n)}{1+z/n} \tag{8.27}$$

It follows that $\Gamma(z)$ is a meromorphic function without zeros and has simple poles precisely at all the non positive integers. Due to the factor $\exp(-\gamma z)$ the function has the following easily verified *Riemann's Functional relation:*

$$\Gamma(z+1) = z\Gamma(z). \tag{8.28}$$

Also, from (8.23) we obtain the product decomposition

$$\frac{\pi}{\sin \pi z} = \Gamma(z)\Gamma(1-z). \tag{8.29}$$

The choice of the constant γ, can be justified in the property that $\Gamma(1) = 1$. Repeated application of (8.28) now yields:

$$\Gamma(n) = (n-1)!, \qquad \forall\ n \in \mathbb{N}. \tag{8.30}$$

Thus one can view Γ as a generalization of the factorial. Its importance in Statistics, perhaps, stems out of this fact.

Exercise 8.7

1. Show that $\Gamma(1/2) = \sqrt{\pi}$.

2. Show that the Gaussian psi-function $\Psi(z) = \dfrac{\Gamma'(z)}{\Gamma(z)}$ is a meromorphic function in $\mathbb{C}$ with simple poles at $z = 0, -1, \ldots, -n, \ldots$ and the residues given by $Res_{-n}(\Psi) = -1$, for all $n \geq 0$. Also prove that

   (a)
$$\frac{\Gamma'(z)}{\Gamma(z)} = -\gamma - \frac{1}{z} + \sum_{1}^{\infty} \frac{z}{n(n+z)}. \tag{8.31}$$

   (b) $\Psi(1) = -\gamma$,

   (c) $\Psi(z+1) - \Psi(z) = \dfrac{1}{z}$

   (d) $\Psi(z) - \Psi(1-z) = -\pi \cot \pi z$.

3. Show that the second derivative of $\ln \Gamma(z)$ is given by
$$\Psi'(z) = \frac{d}{dz}\left(\frac{\Gamma'(z)}{\Gamma(z)}\right) = \sum_{n=0}^{\infty} \frac{1}{(z+n)^2}. \tag{8.32}$$

4. Using the fact that $\Gamma(z)\Gamma(z+1/2)$ and $\Gamma(2z)$ have the same set of poles, we can write
$$\Gamma(z)\Gamma(z+1/2) = \exp(\alpha(z))\Gamma(2z) \tag{8.33}$$
   for some entire function $\alpha(z)$. By using 8.32, show that α is a linear function $\alpha(z) = az + b$. Then using the values of $\Gamma(1/2)$, $\Gamma(1)$ and $\Gamma(3/2)$ solve for a and b :
$$a = -2\ln 2 : \; b = 1/2 \ln \pi + \ln 2.$$
   Thus obtain the **Legendre duplication formula**:
$$\boxed{\boxed{\sqrt{\pi}\Gamma(2z) = 2^{2z-1}\Gamma(z)\Gamma(z+1/2).}} \tag{8.34}$$

8.8 Stirling's Formula

This section will be devoted for deriving the celebrated **Stirling's Formula** for the gamma function:
$$\Gamma(z) = \sqrt{2\pi} z^{z-1/2} e^{-z} e^{J(z)}, \quad \Re(z) > 0 \tag{8.35}$$

where,

$$J(z) := \frac{1}{\pi}\int_0^\infty \frac{z}{\eta^2+z^2}\ln\frac{1}{1-e^{2\pi\eta}}d\eta, \tag{8.36}$$

for all z in the right-half plane $\mathbb{G} = \{z : \Re z > 0\}$. The proof that we have adopted here follows the presentation of the same in [A] which has been attributed to Lindelöff. It gives us a very good opportunity to use our skill in computing the residues in a non trivial way. Consider the summation formula (8.32) for the derivative of the Gaussian psi-function. We would like to represent the rhs of this formula by an integral. Toward this end we seek a good function with residues $\dfrac{1}{(z+n)^2}$ at integral points $z = n$, and our choice falls on

$$\psi(w) = \frac{\pi \cot \pi w}{(z+w)^2} \tag{8.37}$$

(Do not confuse this with the Gaussian psi-function.) Observe that we have fixed $z \in G$ for the time being and treating ψ as a meromorphic function of w on $\mathbb{C}$. Consider the rectangle

$$0 \le \Re w \le n + 1/2; \quad -s \le \Im w \le s$$

for any positive integer n and any positive real number s. The idea is to consider the integrals

$$I_{n,s} = \int_{C_{n,s}} \psi(w)dw$$

of f on the boundary $C_{n,s}$ of the rectangle oriented anti-clockwise and then take the limit as $s \longrightarrow \infty$ and $n \longrightarrow \infty$. By the residue theorem, it follows that the poles $1, 2, \ldots, n$ will contribute successive terms in the rhs of (8.35). The pole $w = 0$ however causes some concern as the contour passes through this point. However, we know that we can add or subtract half of the residue at *zero* as the contour is a straight line segment at this point. In this case, since, the contour will be going around 0 in the anti-clockwise sense, it follows that we have to add $\dfrac{1}{2z^2}$.

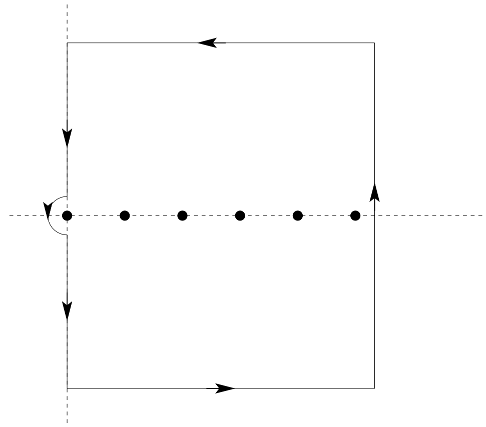


Fig. 40

These considerations yield

$$I_{n,s} = 2\pi\imath\left(\frac{1}{2z^2} + \sum_{k=1}^{n}\frac{1}{(z+k)^2}\right) = 2\pi\imath\left(-\frac{1}{2z^2} + \sum_{k=0}^{n}\frac{1}{(z+k)^2}\right). \tag{8.38}$$

We now observe that $\cot\pi w$ converges uniformly to $\pm\imath$ on the horizontal sides of the rectangle as $s \longrightarrow \infty$. At the same time the denominator of f is seen to tend to ∞. Therefore it follows that

$$\lim_{s\longrightarrow\infty} I_{n,s} = -PV\left(\int_{-\infty}^{\infty}\frac{\pi\cot\pi\imath\eta}{(z+\imath\eta)^2}d(\imath\eta)\right) + PV\left(\int_{-\infty}^{\infty}\frac{\pi\cot\pi(n+1/2+\imath\eta)}{(z+n+1/2+\imath\eta)^2}d(\imath\eta)\right) \tag{8.39}$$

We now show that the second integral in (8.39) vanishes as $n \longrightarrow \infty$. First observe that $|\cot\pi(n+1/2+\imath\eta)| < 1$ for all n and for all η. On the other hand, by using residues we see that,

$$\begin{aligned} PV\left(\int_{-\infty}^{\infty}\frac{d\eta}{|z+n+1/2+\imath\eta|^2}\right) &= PV\left(\int_{-\infty}^{\infty}\frac{d\eta}{(z+n+1/2+\imath\eta)(\bar{z}+n+1/2-\imath\eta)}\right) \\ &= \frac{2\pi\imath}{2n+1+2x} \end{aligned}$$

for large n where $z = x + \imath y$. It follows that, the limit of this second integral is seen to be zero as claimed. Therefore we have,

$$-PV\left(\int_{-\infty}^{\infty}\frac{\pi\cot(\pi\imath\eta)}{(z+\imath\eta)^2}d\eta\right) = 2\pi\left(-\frac{1}{2z^2} + \sum_{k=0}^{\infty}\frac{1}{(z+k)^2}\right). \tag{8.40}$$

Finally,

$$\begin{aligned}
-PV\left(\int_{-\infty}^{\infty}\frac{\pi\cot(\pi\imath\eta)}{(z+\imath\eta)^2}d\eta\right) &= \pi\int_0^{\infty}\cot\pi\imath\eta\left[\frac{1}{(\imath\eta-z)^2}-\frac{1}{(\imath\eta+z)^2}\right]d\eta\\
&= 2\pi\int_0^{\infty}\coth(\pi\eta)\,\frac{2\eta z}{(\eta^2+z^2)^2}d\eta\\
&= 2\pi\int_0^{\infty}\left(1+\frac{2}{e^{2\pi\eta}-1}\right)\frac{2\eta z}{(\eta^2+z^2)^2}d\eta\\
&= 2\pi\left(\frac{1}{z}+\int_0^{\infty}\frac{4\eta z}{(\eta^2+z^2)^2}\frac{d\eta}{e^{2\pi\eta}-1}\right)
\end{aligned}$$

Hence, we have,

$$\frac{d}{dz}\left(\frac{\Gamma'(z)}{\Gamma(z)}\right)=\frac{1}{z}+\frac{1}{2z^2}+\int_0^{\infty}\frac{4\eta z}{(\eta^2+z^2)^2}\frac{d\eta}{e^{2\pi\eta}-1} \tag{8.41}$$

The first stage of our task is over. The next stage is to integrate the relation (8.41) twice w.r.t z to obtain the desired result. We observe that the integral in (8.41) allows integration w.r.t. z once at least, because the integral so obtained

$$\int_0^{\infty}\frac{2\eta}{\eta^2+z^2}\frac{d\eta}{e^{2\pi\eta}-1}$$

is uniformly convergent on compact sets in the right-half plane G and hence differentiation under the integral sign is valid. Therefore, we obtain

$$\frac{\Gamma'(z)}{\Gamma(z)}=C+\text{Ln } z-\frac{1}{2z}+\int_0^{\infty}\frac{2\eta}{\eta^2+z^2}\frac{d\eta}{e^{2\pi\eta}-1} \tag{8.42}$$

Here $Ln\,z$ denotes the principle branch of the logarithm and C is an integration constant to be determined later on. In the next stage of integration w.r.t. z, we see that the multiple valued function $\arctan(z/\eta)$ appears, involving both z and η. In order to avoid this, we first carry out integration by parts once:

$$\int_0^{\infty}\frac{2\eta}{\eta^2+z^2}\frac{d\eta}{e^{2\pi\eta}-1}=\frac{1}{\pi}\int_0^{\infty}\frac{z^2-\eta^2}{(\eta^2+z^2)^2}\ln(1-e^{-2\pi\eta})d\eta \tag{8.43}$$

Observe that here the logarithmic term is a real valued function. Now for similar reasons of compact convergence as above, we can carry out integration w.r.t. z to obtain

$$\ln\Gamma(z)=C_2+C_1z+\left(z-\frac{1}{2}\right)\ln z+\frac{1}{\pi}\int_0^{\infty}\frac{z}{\eta^2+z^2}\ln\left(\frac{1}{1-e^{2\pi\eta}}\right)d\eta. \tag{8.44}$$

Denoting the last integral by $J(z)$, we have,

$$\ln\Gamma(z)=C_2+C_1z+\left(z-\frac{1}{2}\right)\ln z+J(z) \tag{8.45}$$

It remains to compute the values of the constants. For this purpose, we would like to take limits as $z \longrightarrow \infty$. This requires us to know the behavior of $J(z)$as $z \longrightarrow \infty$. We must be careful here not to allow z to approach the imaginary axis. To be precise, we shall find out the limit of $J(z)$ as $z \longrightarrow \infty$ only inside

$$\mathbb{G}_\lambda := \{z \ : \ \Re(z) > \lambda\}$$

for some $\lambda > 0$. For this purpose, write πJ as a sum of two integrals

$$\pi J(z) = \int_0^{|z|/2} \quad + \int_{|z|/2}^\infty = J_1(z) + J_2(z).$$

In the first integral $|\eta^2 + z^2| \geq |z|^2 - \dfrac{|z|^2}{4} = \dfrac{3}{4}|z|^2$. Therefore,

$$|J_1(z)| \leq \frac{4}{3|z|}\int_0^{|z|/2} \ln \frac{1}{1-e^{-2\pi\eta}} d\eta \leq \frac{4}{3|z|}\int_0^{\infty} \ln \frac{1}{1-e^{-2\pi\eta}} d\eta.$$

In the second integral, use the fact that $|\eta^2 + z^2| = |\imath\eta + z||\imath\eta - z| \geq |z||\Re(z) > |z|\lambda$. Hence,

$$|J_2(z)| \leq \frac{1}{\lambda}\int_{|z|/2}^{\infty} \ln \frac{1}{1-e^{-2\pi\eta}} d\eta$$

Hence, as $z \longrightarrow \infty$, it is seen that both J_1 and J_2 tend to zero as required.

We now use the functional relation (8.28) to see that $ln\Gamma(z+1) = \ln z + \ln\Gamma(z)$ and hence from (8.44), we have,

$$C_1 = -\left(z + \frac{1}{2}\right)\ln\left(1 + \frac{1}{z}\right) + J(z) - J(z+1)$$

and taking the limit as $z \longrightarrow \infty$, we get $C_1 = -1$.

In order to evaluate C_2, we use the relation (8.29) for $z = 1/2 + \imath y$. Combining this with 8.45, this gives

$$2C_2 - 1 - \ln\pi = -J(1/2+\imath y) - J(1/2-\imath y) - \ln\cosh\pi y + \imath y\left[\ln(1/2-\imath y) - ln(1/2+\imath y)\right]$$

We now take the limit of rhs as $y \longrightarrow \infty$ and use the facts

$$\lim_{y\longrightarrow\infty} \frac{\cosh \pi y}{e^{\pi y}} = \frac{1}{2} \tag{8.46}$$

and

$$\lim_{y\longrightarrow\infty} \imath y\left[\pi\imath + \ln\left(\frac{1/2-\imath y}{1/2+\imath y}\right)\right] = -1. \tag{8.47}$$

(See Ex.4.6.7.) By adding and subtracting πy, on the rhs, we get,

$$\begin{aligned} 2C_2 - 1 - \ln \pi &= -J(1/2 + \imath y) - J(1/2 - \imath y) &-& \ln(\cosh \pi y) + \pi y \\ && + & \imath y \left[\pi \imath + \ln \frac{1/2 - \imath y}{1/2 + \imath y} \right] \\ &= -J(1/2 + \imath y) - J(1/2 - \imath y) &-& \ln \frac{\cosh \pi y}{e^{\pi y}} + \imath y \left[\pi \imath + \ln \frac{1/2 - \imath y}{1/2 + \imath y} \right] \end{aligned}$$

Upon taking the limit as $y \longrightarrow \infty$ we get,

$$2C_2 - 1 - \ln \pi = \ln 2 - 1$$

and hence $C_2 = \dfrac{1}{2} \ln 2\pi$. This establishes the formula (8.35).

As an important corollary, we can deduce

$$\boxed{\Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\, dt} \tag{8.48}$$

valid in the right-half plane $\mathbb{G}$.

Indeed quite often formula (8.48) goes under the name of Stirling's formula and one can perhaps prove it independently. . In many elementary expositions this is taken as the definition of the Gamma function.

Again, we first observe that the integral on the rhs is uniformly convergent on $\{z : \Re z > \epsilon\}$ for every $\epsilon > 0$. Hence it defines an analytic function in $\mathbb{G}$. We shall denote this function momentarily by $F(z)$. Carrying out integration by parts once, we observe that

$$F(z+1) = \int_0^\infty e^{-t} t^z dt = z \int_0^\infty e^{-t} t^{z-1} dt = zF(z). \tag{8.49}$$

Therefore it follows that the holomorphic function $h(z) = \dfrac{F(z)}{\Gamma(z)}$ is periodic with period 1, i.e.,

$$h(z+1) = h(z).$$

Now observe that the strip $1 \leq \Re z \leq 2$ maps surjectively onto $\mathbb{C}^\star$ under the exponential map. Therefore the periodicity of h implies that ψ factors through the map $z \mapsto e^{2\pi \imath z}$ to define an analytic function $g : \mathbb{C}^\star \longrightarrow \mathbb{C}$, i.e. $g(e^{2\pi \imath z}) = \psi(z)$ for all $z \in \mathbb{G}$. (Incidentally, this also shows that h extends to an entire function.) We claim that g has a removable singularity both at 0 and ∞. This will prove that g is a constant function and hence h too is a constant function. Since it is easily seen that $h(1) = 1$, it would follow that $F(z) = \Gamma(z)$ for all $z \in \mathbb{G}$, as required.

In order to prove that g is a constant function, we recall that it is enough to prove that

$$\lim_{w \longrightarrow 0} |wg(w)| = 0 \text{ and } \lim_{w \longrightarrow \infty} |g(w)/w| = 0 \tag{8.50}$$

in order to prove that both 0 and ∞ are removable singularities. Observe that the portions of the strip given by $\Im z > r$ are mapped onto punctured disc neighborhoods of 0 in $\mathbb{C}$ by the exponential mapping. This can be expressed by saying that "as $y \longrightarrow +\infty$, $w = e^{2\pi \imath z} \longrightarrow 0$." Thus the first part of (8.50) is equivalent to say that for all $1 \le \Re z \le 2$, we have,

$$\lim_{y \longrightarrow \infty} \left| e^{2\pi\imath(x+\imath y)} \frac{F(x+\imath y)}{\Gamma(x+\imath y)} \right| = 0 \tag{8.51}$$

In order to prove 8.51, we observe that

$$|F(z)| \le \int_0^\infty |e^{-t}t^{z-1}|dt = \int_0^\infty e^{-t}t^{x-1}dt = F(x)$$

is bounded in $1 \le x \le 2$. Similarly, since $J(z) \longrightarrow 0$ as $y \longrightarrow \infty$, it follows that there is a positive constant L such that

$$\left| \frac{\Gamma(z)}{z^{z-1/2}} \right| = \sqrt{2\pi}|e^{-z}||J(z)| \ge L$$

say, for all sufficiently large y. Thus

$$\begin{aligned}
\lim_{w \longrightarrow 0} |wg(w)| &= \lim_{y \longrightarrow \infty} \left| e^{2\pi\imath(x+\imath y)} \frac{F(x+\imath y)}{\Gamma(x+\imath y)} \right| \\
&\le M \lim_{y \longrightarrow \infty} \left| \frac{e^{-2\pi y}}{z^{z-1/2}} \right| \\
&\le M' \lim_{y \longrightarrow \infty} \exp(-2\pi y - (x - 1/2)\ln|z| + y \arg z) = 0
\end{aligned}$$

since, $|\arg z| < \pi/2$, for $z \in G$. For similar reasons one can show that $\lim_{w \longrightarrow \infty} \left| \frac{g(w)}{w} \right| = 0$. This completes the proof of the formula (8.48).

Exercise 8.8 Prove (8.46) and (8.47).

Remark 8.8.1 For a lucid account on Gamma functions, see [Ar].

8.9 Extension of Zeta Function

In this section, we extend the zeta function to the entire plane, as a meromorphic function with a simple pole at $z = 1$. On the way, we shall establish the important Riemann's functional relation for $\zeta(z)$. One idea is to express the product $(z-1)\zeta(z)\Gamma(z)$ over $\Re z > 0$ conveniently so that it allows an analytic extension $f(z)$ to the whole plane. Then we could simply take $\zeta(z) = f(z)/\Gamma(z)$. However, due to some technical reason which will become clear to you soon, we actually modify the function $\zeta(z)\Gamma(z)$ additively rather than by multiplying by $z - 1$.

We begin with the Stirling formula

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \tag{8.52}$$

which is valid in the right-half plane $\Re(z) > 0$. Putting $t = nu$, we get,

$$\Gamma(z) = n^z \int_0^\infty e^{-nu} u^{z-1} du \tag{8.53}$$

valid for all integers $n \geq 1$. Thus, dividing out by n^z and taking the summation over $n \geq 1$, we obtain

$$\Gamma(z)\zeta(z) = \sum_1^\infty \int_0^\infty e^{-nu} u^{z-1} du \tag{8.54}$$

If we were allowed to interchange the summation and integration sign on the rhs of (8.54), we immediately recognize that, this would give us a nice expression for the lhs for

$$\sum_1^\infty e^{-nu} = \frac{1}{e^u - 1}, \quad u > 0, \tag{8.55}$$

Indeed, this is precisely the case. First of all, the summation (8.55) is compactly convergent. Also the integral involved are compactly convergent (at both the end points). This means, first that term-by-term integration is possible:

$$\int_\epsilon^R \frac{u^{z-1}}{e^u - 1} du = \sum_1^\infty \int_\epsilon^R e^{-nu} u^{z-1} du$$

and then we can take the limit as $\epsilon \longrightarrow 0$ and $R \longrightarrow \infty$:

$$\int_0^\infty \frac{u^{z-1}}{e^u - 1} du = \sum_1^\infty \int_0^\infty e^{-nu} u^{z-1} du$$

which when combined with (8.55) gives us:

$$\Gamma(z)\zeta(z) = \int_0^\infty \frac{u^{z-1}}{e^u - 1} du \tag{8.56}$$

valid for $\Re(z) > 1$.

In the first stage, we extend $\Gamma(z)\zeta(z)$ to the part $\Re(z) > 0$. For this purpose, we will consider the rhs of (8.56) and try to rewrite it so that it makes sense in the right-half plane, $\Re(z) > 0$. The idea is then to define $\zeta(z)$ by such an expression in the strip $0 < \Re z \leq 1$, (after dividing out by $\Gamma(z)$). However, observe that we expect ζ to have a simple pole at $z = 1$, and hence we should exclude $z = 1$, from the right-half plane. Alternatively, we should add a suitable multiple of $\dfrac{1}{z-1}$ to the rhs and only then try to extend it to the entire right-half plane and this is precisely what we are going to do. This leads us to examine the polarity of the integrand on the rhs. of (8.56). So, consider the Laurent expansion

$$\frac{1}{e^u - 1} = \frac{1}{u} - \frac{1}{2} + \sum_{n=1}^{\infty} a_n u^n \tag{8.57}$$

We see that $\dfrac{1}{e^u - 1} - \dfrac{1}{u}$ remains bounded in a neighborhood of 0 and hence the integral

$$\int_0^1 \left(\frac{1}{e^u - 1} - \frac{1}{u} \right) u^{z-1} du$$

converges compactly on $\Re z > 0$. Therefore we can rewrite (8.56) in the form,

$$\zeta(z)\Gamma(z) = \int_0^1 \left(\frac{1}{e^u - 1} - \frac{1}{u} \right) u^{z-1} du + \frac{1}{z-1} + \int_1^\infty \frac{u^{z-1}}{e^u - 1} du \tag{8.58}$$

valid in $\Re(z) > 1$. But then the rhs of 8.58 makes sense in the whole of right-half plane as a meromorphic function in with a simple pole at $z = 1$ and has residue $= 1$. We define $\Gamma(z)\zeta(z)$ by equating it with the rhs of 8.58 for $\Re(z) > 0$. The first stage of our task is over.

In the next stage, we shall rewrite (8.58) in such a way that it makes sense in the strip $-1 < \Re z < 1$. For this all we have to do is to observe that for $0 < \Re z < 1$, we have,

$$\frac{1}{z-1} = -\int_1^\infty u^{z-2} du$$

and hence (8.58) becomes

$$\zeta(z)\Gamma(z) = \int_0^\infty \left(\frac{1}{e^u - 1} - \frac{1}{u} \right) u^{z-1} du \tag{8.59}$$

in the strip $0 < \Re z < 1$. We again use (8.57) to see that the function

$$\frac{1}{e^u - 1} - \frac{1}{u} + \frac{1}{2} \leq cu, \ 0 \leq u \leq 1$$

for some constant c and hence,

$$\int_0^1 \left(\frac{1}{e^u - 1} - \frac{1}{u} + \frac{1}{2} \right) u^{z-1} du$$

is uniformly convergent on compact subsets of $\Re z > -1$. Similarly, one can see that

$$\int_1^\infty \left(\frac{1}{e^u - 1} - \frac{1}{u} \right) u^{z-1} du$$

is also uniformly convergent on compact subsets of $\Re(z) > 1$. Hence, we can write (8.59) in the form:

$$\zeta(z)\Gamma(z) = \int_0^1 \left(\frac{1}{e^u - 1} - \frac{1}{u} + \frac{1}{2} \right) u^{z-1} du - \frac{1}{2z} + \int_1^\infty \left(\frac{1}{e^u - 1} - \frac{1}{u} \right) u^{z-1} du \tag{8.60}$$

Since both the integrals on the rhs of (8.60) are compactly convergent in the strip $-1 < \Re z < 1$, we can take rhs divided by $\Gamma(z)$ as the definition of ζ in this strip. It appears as though at $z = 0$ we have some trouble because of the term $-\dfrac{1}{2z}$. However, we observe that Γ also has a simple pole at $z = 0$. Thus ζ has a removable singularity at $z = 0$ and its value at $z = 0$ can be defined by taking the appropriate limit. The second stage of our task is over.

We shall now rewrite (8.60), in the strip $-1 < \Re z < 0$, in such a way that it will make sense in the whole of the left -half plane, $\Re z < 0$. For this we first observe that

$$\frac{1}{z} = -\int_1^\infty u^{z-1} du.$$

Therefore from (8.60), we obtain,

$$\zeta(z)\Gamma(z) = \int_0^\infty \left(\frac{1}{e^u - 1} - \frac{1}{u} + \frac{1}{2} \right) u^{z-1} du \tag{8.61}$$

On the other hand

$$\begin{aligned}\frac{1}{e^u-1}+\frac{1}{2} &= \frac{1}{2}\frac{e^u+1}{e^u-1}\\ &= \frac{\imath}{2}\cot(\imath u/2)\\ &= \frac{\imath}{2}\left[\frac{2}{\imath u}-4\imath u\sum_1^\infty\frac{1}{u^2+4n^2\pi^2}\right]\\ &= \frac{1}{u}+2u\sum_1^\infty\frac{1}{u^2+4n^2\pi^2}\end{aligned}$$

Therefore,

$$\Gamma(z)\zeta(z)=2\int_0^\infty\left(\sum_{n=1}^\infty\frac{1}{u^2+4\pi^2n^2}\right)u^z du \tag{8.62}$$

For reasons similar to the one used above in interchanging the order of summation and integration, it follows that

$$\Gamma(z)\zeta(z)=2\sum_{n=1}^\infty\int_0^\infty\frac{u^z}{u^2+4\pi^2n^2}du \tag{8.63}$$

Putting $u=2\pi nt$, we get,

$$\Gamma(z)\zeta(z)=2\sum_{n=1}^\infty(2\pi n)^{z-1}\int_0^\infty\frac{t^z}{t^2+1}dt$$

This is rewritten as

$$\Gamma(z)\zeta(z)=2(2\pi)^{z-1}\zeta(1-z)\int_0^\infty\frac{t^z}{t^2+1}dt \tag{8.64}$$

valid in the strip $-1<\Re z<0$. Now from (8.29) in section 7, it follows that

$$\zeta(z)=2(2\pi)^{z-1}\zeta(1-z)\Gamma(1-z)\sin(\pi z)\,\frac{1}{\pi}\int_0^\infty\frac{t^z}{t^2+1}dt. \tag{8.65}$$

Once more, we shall appeal to calculus of residues to obtain the formula (see Exercise 1 of section 6.5)

$$\frac{1}{\pi}\int_0^\infty\frac{t^{-x}}{1+t}dt=\operatorname{cosec}\pi x,\quad 0<x<1. \tag{8.66}$$

It is seen easily that for $-1<x<0$ (by substituting $t^2=u$)

$$\frac{1}{\pi}\int_0^\infty\frac{t^x}{t^2+1}dt=\frac{1}{2\pi}\int_0^\infty\frac{u^{(x-1)/2}}{u+1}du=\frac{1}{2}\operatorname{cosec}\left[\pi(1-x)/2\right]$$

and since, both the sides define holomorphic functions in the strip $-1 < \Re z < 0$ when we replace x by z, it follows that the above equality is valid in the entire of the above strip.

Combining this with (8.66) we obtain the *Riemann's Functional Relation:*

$$\boxed{\zeta(z) = 2(2\pi)^{z-1}\Gamma(1-z)\zeta(1-z)\sin(\pi z/2).} \tag{8.67}$$

The validity of this relation is proved in the strip $-1 < \Re z < 0$ so far. We notice that the rhs of this relation is holomorphic in the entire of the left-half plane $\Re z < 0$. Hence we can and do use this to define the zeta function in the left-half plane. This completes the task of extending the definition of zeta function to the entire plane as required.

We now have two holomorphic functions on either side of 8.67 agreeing on a non empty on set of the domain $\mathbb{C} \setminus \{1\}$. Therefore the identity is valid in the whole of $\mathbb{C} \setminus \{1\}$.

Exercise 8.9

1. Show that $(1-z)^{-\alpha} = \sum_0^\infty \frac{\Gamma(n+\alpha)}{n!\Gamma(\alpha)} z^n, \quad z \in \mathbb{D}, \alpha > 0.$

2. Show that
$$\lim_{z\to\infty} \frac{\Gamma(z+1)}{z/e)^z\sqrt{2\pi z}} = 1$$
where the limit is taken remaining within the right-half plane $\Re(z) > 0$. This gives an approximate expression for $(n+1)!$ for large n and is also known as Stirling's formula.

3. Show that $-2, -4, \ldots, -2n, \ldots$ are simple zeros of $\zeta(z)$. (These are called *trivial zeros* of $\zeta(z)$.)

4. Show that $\zeta(z)$ has no zeros outside the strip $0 \leq \Re(z) \leq 1$ other than the trivial zeros.

   [This strip is called the *critical strip.* One of the most celebrated problems in mathematics is following conjecture of Riemann:

   All the nontrivial zeros of ζ are on the line $x = \frac{1}{2}$.

   It has remained an open problem even today even after 150 years and goes under the name Riemann-hypothesis. It is known that there are no zeros of ζ on the lines $x = 0$ and $x = 1$. It is also known that there are infinitely many zeros on the line $x = \dfrac{1}{2}$. The importance of this problem can be gauged by the fact that it is one of the seven Millenium problems with a prize money of 1 million dollars.

5. *Euler's Identity*: Prove the identity

$$\boxed{\zeta(z) = \prod_{n=1}^{\infty}\left(\frac{1}{1-p_n^{-z}}\right), \text{ for } \Re z > 1.} \tag{8.68}$$

Deduce that

$$\sum \frac{1}{p_n} = \infty$$

where p_n denotes the n^{th} prime. [This gives a proof that there are infinitely many primes.]

6. Let $\xi(z) = z(z-1)\pi^{-z/2}\zeta(z)\Gamma(z/2)$. Show that ξ is an entire function and satisfies $\xi(z) = \xi(1-z),\ z \in \mathbb{C}$.

8.10 Normal Families and Equicontinuity

The situation of our interest is a family $\mathcal{F}$ of functions with (or without) any specific properties, defined on a region Ω in $\mathbb{C}$, and taking values in $\mathbb{C}$. We shall give only one application of the theory of normal families here, which indeed reasonably covers the central theme. However, our treatment is at best, skeletal and the reader may consult the books by Hille [H], for more information.

We would like to have families of functions with the Bolzano[4]-Weierstrass type property, viz., each sequence should have a subsequence that is convergent. This should immediately ring a familiar bell in us: perhaps we are dealing with some sort of compactness property of the family of functions. Indeed, this is precisely the case. The space of functions can be given a suitable topology with respect to which compactness will be equivalent to this Bolzano-Weierstrass property. We shall not elaborate this aspect any more. The interested reader may see the chapter on function spaces in [J] or [Ke]. Another closely related concept to that of compactness is the concept of uniform continuity. So, we would like to introduce the concept of *'uniform continuity w.r.t. a family*

[4]Bernard Bolzano (1781-1848) an Austrian mathematician and professor of religious studies is known for his studies in point sets and foundational mathematics.

of functions'. This is known as *'equicontinuity'.* We aim to relate these two concepts and apply them to the study of conformal mappings.

Let Ω be an open subset of $\mathbb{C}$ and $\mathcal{C}(\Omega)$ denote the set of all continuous complex valued functions on Ω. In what follows we shall consider subsets of $\mathcal{F} \subset \mathcal{C}(\Omega)$ and refer to them merely as family of functions on Ω.

Definition 8.10.1 A family $\mathcal{F}$ of functions on Ω is said to be *normal* in Ω, if every sequence $\{f_n\}$ in $\mathcal{F}$ admits a subsequence which converges uniformly on every compact subset of Ω.

Observe that it is not required that the limit function also belongs to $\mathcal{F}$.

Definition 8.10.2 Let $\mathcal{F}$ be a family of functions on Ω. We say, $\mathcal{F}$ is *equicontinuous* on a subset $A \subseteq \Omega$, if for every $\epsilon > 0$ there exists $\delta > 0$, such that $|f(z), f(w)| < \epsilon$ for all $|z - w| < \delta$, for all $z, w \in A$ and $f \in \mathcal{F}$.

Remark 8.10.1 Clearly, if $\mathcal{F}$ is a equicontinuous family of functions then each member of this family is uniformly continuous on A.

Theorem 8.10.1 Arzela- Ascoli: *Let $\mathcal{F}$ be a family of continuous functions on Ω. Then $\mathcal{F}$ is normal iff*
(1) *$\mathcal{F}$ is equicontinuous on every compact subset of Ω, and*
(2) *for every fixed $z \in \Omega$, the set $A_z = \{f(z) \ : \ f \in \mathcal{F}\}$ is bounded.*

Proof: Let $\mathcal{F}$ be normal. Suppose (1) does not hold. This means that there is a compact $K \subset \Omega$ on which $\mathcal{F}$ is not equicontinuous. This, in turn, means that there is $\epsilon > 0$ such that for every n we have some $f_n \in \mathcal{F}, z_n, w_n \in K$ such that $|z_n - w_n| < 1/n$ and $|f_n(z_n) - f_n(w_n)| \geq \epsilon$. Passing to subsequences we may simply assume that $z_n \to z, w_n \to w$ and f_n converges uniformly on every compact subset of Ω to a function f. Hence for large n we have $|f_n(p) - f(p)| < \epsilon/4$ for all $p \in K$. Clearly f is continuous. Since $|w_n - z_n| < 1/n$ it follows that $z = w \in K$. By continuity of f, it follows that for large n, we have $|f(z_n) - f(z)| < \epsilon/4$ and similarly $|f(w_n) - f(w)| < \epsilon/4$ which leads to a contradiction that $|f_n(w_n) - f_n(z_n)| < \epsilon$ for large n.

To prove (2) let us show that $\overline{A_z}$ is compact. So let w_n be a sequence in $\overline{A_z}$. Then there exist $f_n \in \mathcal{F}$, such that $d(w_n, f_n(z)) < 1/n$. By normality of $\mathcal{F}$, we have a subsequence f_{n_k} which is convergent. But then $f_{n_k}(z)$ converges to a point $w \in \overline{A_z}$. The sequence w_n also is seen to converge to the same point. Hence $\overline{A_z}$ is compact.

[To prove the converse, we have to invoke the famous '*diagonal process*' due to Cantor. Given a sequence in $\mathcal{F}$, to find a subsequence that is convergent at a given point $z \in \Omega$ is easy from (2). Thus, in principle by the diagonal process we can get a subsequence which is convergent at a given countable subset of Ω. This should, in effect suffice our purpose because of the separability property of $\mathbb{C}$. The rest of the requirements are fulfilled by condition (1). This is only the idea of the proof. Let us now write down the proof carefully.]

Let $\{z_n\}$ be a sequence of points which is dense in Ω. (For this, we can simply take all points in Ω with rational coordinates and enumerate them.) Let $\{n_{1,j}\}$ be a subsequence of $\{n\}$ such that $\lim_j f_{n_{1,j}}(z_1) = w_1$. Now, inductively, having found a sequence $\{n_{i,j}\}$, which is a subsequence of $\{n_{i-1,j}\}$ such that $\lim_j f_{n_{i,j}}(z_i) = w_i$, we find the next sequence with the same property with i replaced by $i+1$. Finally, let $m_j = n_{j,j}$. Then f_{m_j} is a subsequence of f_n such that $\lim_j f_{m_j}(z_i) = w_i$ for all i.

[It is easy to see that this sequence converges to a function f on the set $\{z_n\}$. However, in order to conclude that f is continuous, and also extend it to the whole of Ω, we need to show that the convergence is uniform on each compact subset of Ω. Instead, we shall actually show that the sequence is *uniformly Cauchy* on each compact subset K of Ω. Since the space of continuous functions is complete, that is good enough.]

Given $\epsilon > 0$, by equicontinuity of $\mathcal{F}$, we get $\delta > 0$ such that, for all $f \in \mathcal{F}$, and $z, w \in K$, such that $|z - w| < \delta$, we have, $|f(z) - f(w)| < \epsilon/3$. Since K is compact, we can cover K with finitely many balls $B_1, \ldots, B_r$ of radius $\delta/2$. Since the set A= $\{z_n\}$ is dense, we can pick one point each from $A \cap B_j$, say, without loss of generality, $z_j \in A \cap B_j,\ 1 \leq j \leq r$. By the convergence of the sequences $f_{m_j}(z_k)$, there exists N such that if $i, j \geq N$ then $|f_{m_i}(z_k) - f_{m_j}(z_k)| < \epsilon/3$ for all $1 \leq k \leq r$. Finally, let $z \in K$ be any. Then $z \in B_k$, say. As usual, using triangle inequality now, we get $|f_{m_i}(z) - f_{m_j}(z)| < \epsilon$. This shows that the sequence $f_{m_j}(z)$ is uniformly Cauchy on each compact subset K of Ω. ♠

The following theorem is the only result about the normal families, that we are going to put to use immediately.

Theorem 8.10.2 Montel:[5] *A family $\mathcal{F}$ of holomorphic functions defined in a domain Ω is normal iff it is uniformly bounded on every compact subset of Ω.*

[5]Paul Antoine Aristide Montel (1876-1975), a French mathematician, was a late comer in mathematics. Apart from from his fundamental ideas in normal families he has also investigated the relation between the coefficients of a polynomial and the location of its zeros in the complex plane.

Proof: Suppose $\mathcal{F}$ is normal. Given a compact subset $K \subset \Omega$, we want to find a uniform bound M_1 for all $f \in \mathcal{F}$ over K. Take any $\epsilon > 0$. By equicontinuity, there exist $\delta > 0$, such that for all $z, w \in K$ and $|z - w| < \delta$, we have $|f(z) - f(w)| < \epsilon$. By the compactness of K there exist finitely many balls $B_r(z_j)$ covering K, where $z_1, \ldots, z_k \in K$ and $r = \delta/2$. By condition (2) of the theorem 8.10.1 for normality, there exist M such that $|f(z_j)| < M$, for all $f \in \mathcal{F}$, and for all $1 \leq j \leq k$. Now it is easily verified that $|f(z)| < M + \epsilon =: M_1$ for all $z \in K$ and for all $f \in \mathcal{F}$.

Now let us prove the converse. For $z \in \Omega$ choose $r > 0$, such that the ball $B_{4r}(z) \subset \Omega$. Let C be the boundary circle of $B_{4r}(z)$. Then for $w, w' \in B_{2r}(z)$, we have, by Cauchy's integral formula,

$$\begin{aligned} f(w) - f(w') &= \frac{1}{2\pi\imath}\int_C \left(\frac{1}{\zeta - w} - \frac{1}{\zeta - w'}\right) f(\zeta)d\zeta \\ &= \frac{w - w'}{2\pi\imath}\int_C \frac{f(\zeta)d\zeta}{(\zeta - w)(\zeta - w')}. \end{aligned}$$

Let M be a uniform bound for all $f \in \mathcal{F}$ on C. Then it follows that by M-L inequality that

$$|f(w) - f(w')| \leq \frac{M|w - w'|}{r}. \tag{8.69}$$

Now let K be a compact subset. Cover it with finitely many such balls $B_{r_j}(z_j), j = 1, 2, \ldots, k$. Let M_j be the bound of $|f(z)|$ for all $f \in \mathcal{F}$, on the boundary of $B_{4r_j}(z_j)$ chosen as above. Let $r = min\{r_j\}$, and $M = max\{M_j\}$. For a given $\epsilon > 0$ let $\delta = min\{r, r\epsilon/M\}$. Now let $w, w' \in K$ be such that $|w - w'| < \delta$. Suppose $w \in B_{r_j}(z_j)$. Then $|w' - z_j| < \delta + r_j \leq 2r_j$. Therefore both w, $w' \in B_{2r_j}(z_j)$. Hence from (8.69), it follows that, $|f(w) - f(w')| \leq M_j\delta/r_j \leq M\delta/r \leq \epsilon$. ♠

8.11 Uniformization: Riemann Mapping Theorem

In this section, our aim is to prove the celebrated theorem of Riemann: *any simply connected domain in* $\mathbb{C}$ *is either* $\mathbb{C}$ *or is biholomorphic to the open unit disc* $\mathbb{D}$. There are many proofs of this theorem as every good theorem should have. The original proof due to Riemann contained a few gaps which were filled up later on. However, the first correct proof (due to Koebe[6]) was given in a line of argument quite different from that

[6]Paul Koebe(1882-1945) was a German mathematician whose main contribution is in establishing the general principle of uniformization. He was known for his pompous and chaotic style.

of Riemann. Here, we shall present a slightly different version of Koebe's proof. It may be remarked that there are similar but more complicated (even to state) results about the classification of multi-connected domains. For the proof of these statements, particularly, for domains of connectivity greater than two, perhaps one has to go back to Riemann's line of argument. The proof depends only on results in section 8.1 and 8.9. Indeed, we need only the 'only if' part of Arzela-Ascoli theorem 8.10.1.

As a warm up for conformal mappings, recall that we have a large number of them: the translations, the rotations, the inversion, the dilation and more generally the fractional linear transformations, the exponential function, the branches of logarithm function and root-functions and so on. We now introduce the so called **Koebe transformation** which is an extremely ingenious combination of a few of the above type of conformal mappings:

Let $0 < a < 1,\ 0 \leq \theta \leq 2\pi$. Define

$$R_\theta(z) = (\exp \imath\theta)z; \qquad \tau_a(z) = \frac{a - z}{1 - az}. \tag{8.70}$$

Observe both R and τ are biholomorphic mappings of $\mathbb{D}$ onto itself. Let B be any simply connected region in $\mathbb{D}$ and let $0 \notin B$. Define

$$q_B(z) = \sqrt{z},$$

where q_B is one of the two square-root functions which exist on B because of the simply connectivity of B and $0 \notin B$. Clearly then q_B is a holomorphic injective mapping. Now the following lemma is easy to prove. The transformations κ defined below are called *Koebe transformations.*

Lemma 8.11.1 Let A be a simply connected region in $\mathbb{D}$, $0 \in A$ and $w = a \exp \imath\theta,\ 0 < a < 1$, be a point of $\mathbb{D} \setminus A$. Put $B = \tau_a \circ R_{-\theta}(A)$ and $\kappa := \kappa(A, w) := R_\theta \circ \tau_{\sqrt{a}} \circ q_B \circ \tau_a \circ R_{-\theta}$. Then κ is a biholomorphic mapping of A into $\mathbb{D}$ and we have

$$\kappa(0) = 0; \quad \kappa'(0) = \frac{1 + a}{2\sqrt{a}} > 1.$$

Proof: Since each of the five mappings of which κ is the composite is biholomorphic, κ is also biholomorphic. The values $\kappa(0), \kappa'(0)$ are easily computed. ♠

Remark 8.11.1 It is interesting to note that if λ denotes the inverse of κ, then λ is actually defined on all of $\mathbb{D}$ and maps $\mathbb{D}$ into itself. Obviously, it is not a rotation and hence by Schwarz's lemma, it follows that $|\lambda'(0)| < 1$. Therefore, we can conclude that $|\kappa'(0)| > 1$ without any computation. We are now ready to prove:

Theorem 8.11.1 Riemann Mapping Theorem: *Let Ω be a simply connected domain in $\mathbb{C}$ not equal to the whole plane. Then given any $z_0 \in \Omega$ there exists a unique biholomorphic mapping $f : \Omega \longrightarrow \mathbb{D}$ such that $f(z_0) = 0$ and $f'(z_0)$ is a positive real number.*

Step I: We shall first dispose of the uniqueness part. If g is another such biholomorphic mapping, then $h = g \circ f^{-1}$ is a biholomorphic mapping of $\mathbb{D}$ onto itself such that $h(0) = 0$ and $h'(0) > 0$. Hence, by Schwarz's lemma it follows that $h(z) = z$ for all $z \in \mathbb{D}$. [For, if h^{-1} denotes its inverse, then by the chain rule it follows that $(h^{-1})'(0) = (h'(0))^{-1}$. Since both $|h'(0)|$ and $|(h^{-1})'(0)|$ are less than or equal to 1, it follows that they must be equal to one. Hence, by the 'equality' part of the Schwarz's lemma, $h(z) = cz$ for some c with $|c| = 1$. Since $h'(0) > 0$ the conclusion follows.] This is the same as saying, $f(z) = g(z)$ for all $z \in \Omega$.

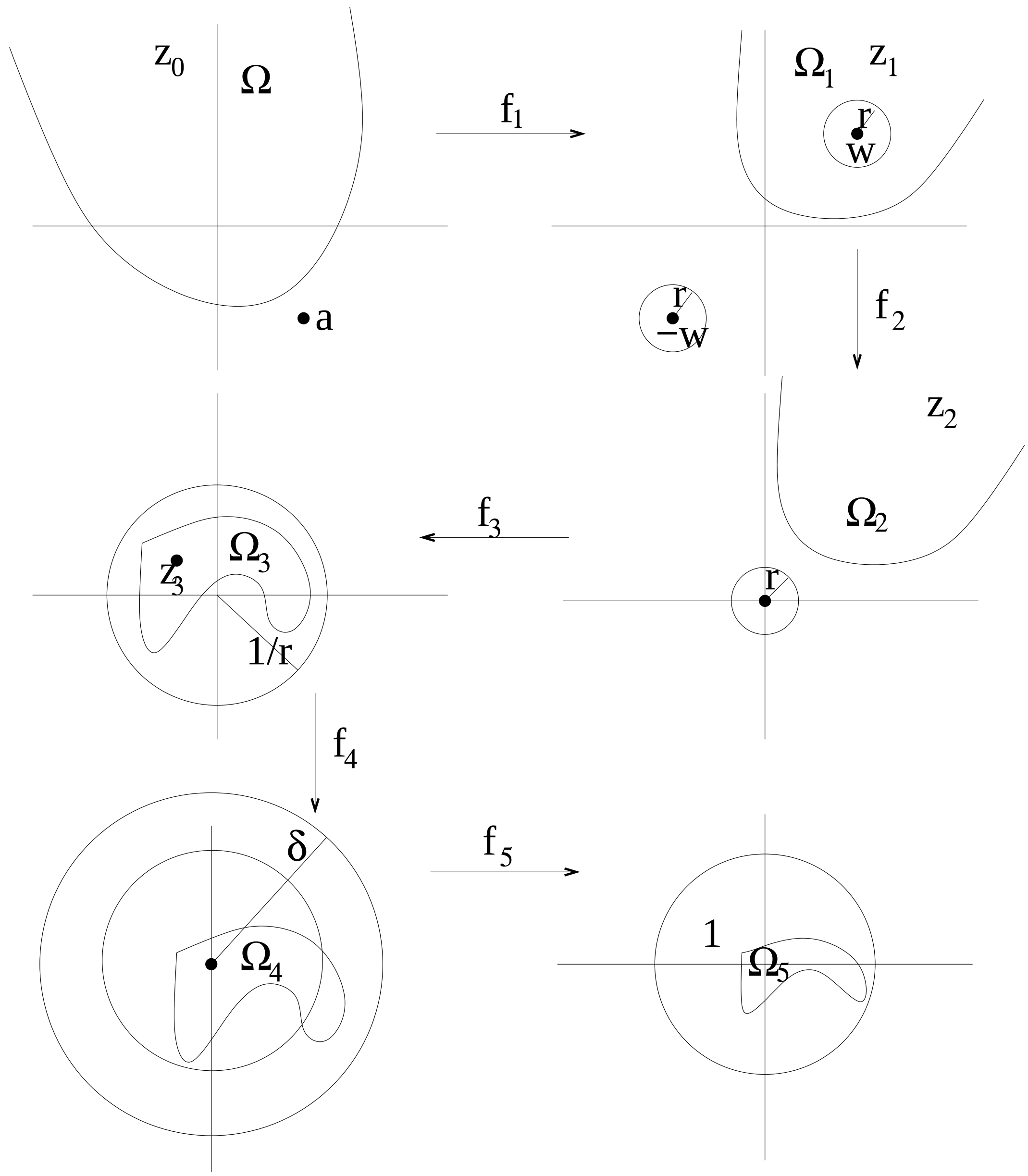


Fig. 41

Step II : Next we will reduce the problem to the case when $\Omega \subset \mathbb{D}$, and $z_0 = 0$. (See Fig. 41.) Given any simply connected proper subset Ω of $\mathbb{C}$, let $a \in \mathbb{C} \setminus \Omega$. Since Ω is simply connected it follows that we have a well defined branch $f_1(z) = \sqrt{z-a}$ on Ω. Put $f_1(\Omega) = \Omega_1$. Observe that f_1 maps Ω onto Ω_1 biholomorphically. Moreover, $\Omega_1 \cap (-\Omega_1) = \emptyset$. Choose w and $r > 0$ are such that the ball $B_r(-w) \subset -\Omega_1$. Then it follows that $B_r(-w) \cap \Omega_1 = \emptyset$. Let f_2 be the translation $f_2(z) = z + w$, $\Omega_2 = f_2(\Omega_1)$. Then we have, f_2 is a biholomorphic mapping and $B_r(0) \cap \Omega_2 = \emptyset$. Now consider the inversion, $f_3(z) = 1/z$. Then f_3 is a biholomorphic mapping of Ω_2 onto $\Omega_3 = f_3(\Omega_2)$. Observe that Ω_3 is now a bounded region being contained in $B_{1/r}(0)$.

Let $z_3 = f_3 \circ f_2 \circ f_1(z_0)$. Now consider another translation, $f_4(z) = z - z_3$. Let δ be such that $|z| < \delta$ for all $z \in f_4(\Omega_3) =: \Omega_4$. Let $f_5(z) = z/\delta$. Then clearly, both f_4 and f_5 are biholomorphic. If $\Omega_5 = f_5(\Omega_4)$, we see that $\Omega_5 \subseteq \mathbb{D}$. Moreover, $f_5 \circ f_4(z_3) = 0$. Thus if we set $g = f_5 \circ \cdots \circ f_1$ then we have a biholomorphic mapping $g : \Omega \longrightarrow \Omega_5 \subseteq \mathbb{D}$ such that $g(z_0) = 0$. If $g'(z_0) = a \exp \imath\alpha$, we then perform a rotation of the unit disc through an angle $-\alpha$, i.e., let $R(z) = z \exp(-\imath\alpha)$, and let $g_1 = R \circ g$. Then $g_1'(z_0) = a > 0$.

Now, by replacing Ω by $g_1(\Omega)$ we can and will assume that $\Omega \subset \mathbb{D}$ is a simply connected region and $z_0 = 0$.

Step III: We are seeking a biholomorphic mapping f of Ω onto $\mathbb{D}$ such that $f(0) = 0$, $f'(0) > 0$. Let now $\mathcal{F}$ be the family of all injective holomorphic mappings $h : \Omega \longrightarrow \mathbb{D}$, such that $h(0) = 0$ and $h'(0) > 0$. This family is clearly non empty since it contains the inclusion mapping. Since all the members take values inside the unit disc, it follows that $\mathcal{F}$ is a normal family. We want to find $f \in \mathcal{F}$ for which $f'(0)$ is maximal and then claim that this f should also be a surjective mapping and hence a biholomorphic mapping.

Let M be the supremum of $\{h'(0) \ : \ h \in \mathcal{F}\}$. There exists a sequence h_n in $\mathcal{F}$ such that $h_n'(0) \longrightarrow M$. Since $\mathcal{F}$ is normal, there is a subsequence h_{n_k} which is $\rho-$convergent to say, f. Clearly then $f : \Omega \longrightarrow \mathbb{D}$ is holomorphic on Ω and $f'(0) = M$. Also $M > 0$ and hence f is not a constant. Since each h_n is injective, by a corollary to Hurwitz theorem (see theorem 8.1.5), it follows that so is f.

It remains only to prove that $f(\Omega) = \mathbb{D}$ and this is where we use the so called Koebe transformations. If there exists $w \in \mathbb{D} \setminus f(\Omega)$, take $\kappa := \kappa(f(\Omega), w)$ as defined in (8.70). Then the composition $\phi = \kappa \circ f$, is also a member of $\mathcal{F}$. Also, by chain rule, we see that $\phi'(0) = \kappa'(0) f'(0) > M$. This contradicts the maximality of M. Therefore $f(\Omega) = \mathbb{D}$, and the proof of RMT is complete. ♠

Chapter 9

Dirichlet's Problem

Previously, in section 4.9, we have touched upon this subject a little bit and have promised to do a bit more later. In this chapter we shall study Perron's solution of this problem. As a major application, we shall present another proof of Riemann-Mapping Theorem. The method employed goes much beyond this. It can be fruitfully adopted in the classification of all multi-connected domains in $\mathbb{C}$ as well, which we shall only indicate.

9.1 Mean Value Property

As promised in 4.9, it is time to re-visit the harmonic functions. We shall treat them as independently as possible rather than as a part of holomorphic function. As a first step, let us reprove the mean value property in a slightly general context.

Theorem 9.1.1 *Let $\Omega = A(Z_0; R_0, R_1)$ be an annular domain with center z_0 and $u : \Omega \to \mathbb{R}$ be a harmonic function. Then for all $R_0 < r < R_1$, the arithmetic mean of u on the circle $|z - z_0| = r$ is a linear function of $\ln r$, i.e., there exist real numbers α, β such that*

$$\frac{1}{2\pi}\int_0^{2\pi} u(re^{\imath\theta})d\theta = \alpha \ln r + \beta, \quad R_0 < r < R_1. \tag{9.1}$$

Proof: Without loss of generality, we may assume that $z_0 = 0$. If we set

$$f(r) = \frac{1}{2\pi}\int_0^{2\pi} u(re^{\imath\theta})d\theta,$$

then clearly f is a real valued smooth function on the open interval (R_0, R_1). It suffices to show that $\frac{df}{dr} = \frac{\alpha}{r}$, where α is a constant. Differentiating under the integral sign, we get, $\dfrac{df}{dr} = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} \dfrac{\partial u}{\partial r}\, d\theta$. Therefore, our task is to show that

$$\int_0^{2\pi} r\frac{\partial u}{\partial r} d\theta = \text{a constant} \tag{9.2}$$

in the interval (R_0, R_1). Once again we use differentiation under the integral sign w.r.t. to r. From the polar coordinate form of Laplace's equation (4.45), it follows that the derivative of the integrand in (9.2) equals $-\dfrac{1}{r}\dfrac{\partial^2 u}{\partial \theta^2}$. Therefore we have to prove that

$$\int_{|z|=r} \frac{1}{r}\frac{\partial^2 u}{\partial \theta^2}\, d\theta = 0. \tag{9.3}$$

But for a fixed r, $\dfrac{\partial^2 u}{\partial \theta^2}\, d\theta = d(u_\theta)$ and hence integral (9.3) is equal to $\frac{1}{r}[u_\theta(re^{2\pi\imath\theta}) - u_\theta(r)] = 0$. ♠

Now suppose that Ω is a disc with center z_0 and u is harmonic in $\Omega \setminus \{z_0\}$ and continuous at z_0. Then taking limit as $r \to 0$ the formula (9.1) shows that the constant $\alpha = 0$. Using Mean Value Theorem of integral calculus, and continuity of u, we now conclude that $\beta = u(z_0)$.

Thus we have:

Theorem 9.1.2 Mean Value Property : *Let u be harmonic in a domain Ω and $B_r(z_0) \subset \Omega$. Then*

$$u(z_0) = \frac{1}{2\pi}\int_0^{2\pi} u(z_0 + re^{\imath t})\, dt. \tag{9.4}$$

Remark 9.1.1 Recall that we have derived (9.4) directly from Gauss Mean Value Theorem. The novelty here is in the approach as well as under seemingly weaker hypothesis, viz, u is harmonic in $\Omega \setminus \{z_0\}$ and bounded around z_0, (9.4) gives the value for u to be extended continuously. Analogous to holomorphic functions, we can describe this situation by saying that z_0 is a removable singularity of u. Indeed let us prove:

Theorem 9.1.3 *Let u be harmonic in $B_r(z_0) \setminus \{z_0\}$. Then the following conditions are equivalent.*
(i) *u is bounded in a closed neighborhood of z_0.*
(ii) *z_0 is a removable singularity of u, i.e., u can be extended to a harmonic function on the whole of $B_r(z_0)$.*

Proof: The implication (ii)$\Longrightarrow$(i) is obvious. To see (i) $\Longrightarrow$ (ii), let α and β be as in theorem 9.1.1. It follows that $\alpha = 0$ for otherwise, RHS of (9.1) becomes unbounded near 0 which is absurd. As seen before one expects that if we define $u(z_0) = \beta$ then it will be continuous at z_0 as well.

We claim that u is actually harmonic at z_0. For this, we go back to holomorphic functions, viz., we shall produce a holomorphic function f on $B_r(z_0) \setminus \{z_0\}$ whose real part is equal to u. Then it follows that z_0 is a removable singularity of f and hence u is harmonic at z_0 also. (Compare Exercise 5.3.7)

Once again, for notational simplicity we may asume that $z_0 = 0$. Consider the differential

$$^*du := u_x dy - u_y dx$$

We claim

$$\int_{|z|=r} {}^*du = 0.$$

For when r is a constant, we have

$$dx = -r \sin\theta \, d\theta; \quad dy = r\cos\theta \, d\theta.$$

Therefore $^*du = ru_r d\theta$. But as seen before, $\int_{|z|=r} ru_r \, d\theta = \alpha = 0$.

Therefore $^*du$ is an exact form. (See Exercise 7.4.5.) This means that there exists a function v on the punctured disc such that $dv = {}^*du$. i.e., $v_x = -u_y; v_y = u_x$. Therefore $f := u + \imath v$ satisfies CR equations throughout the punctured disc. Therefore f is holomorphic as required. ♠

Remark 9.1.2 Given an exact differential $du = u_x dx + u_y dy$ the differential $^*du = u_x dy - u_y dx$ is called the *conjugate differential.* What we have seen in the proof of the above theorem amounts to the following: if u is a harmonic function in a domain Ω such that $\int_\gamma {}^*du = 0$ for all closed contours γ in Ω then u is the real part of a holomorphic function on Ω.

Exercise 9.1

1. If f is analytic and does not vanish on a domain, show that $\ln|f(z)|$ is harmonic.

2. Show that a harmonic function $u : \Omega \to \mathbb{R}$ is an open mapping.

3. Let Ω be a region, u be a harmonic function on Ω and Z be the set of zeros of the gradient of u in Ω, i.e., $Z = \{z \in \Omega \ : \ u_x(z) = u_y(z) = 0\}$. Show that Z is a discrete subset of Ω. On the other hand, show that none of the zeros of u can be isolated.

4. Let u be harmonic in the punctured disc $B_r(0) \setminus \{0\}$ and let α and β be as in theorem 9.1.1. In the proof of theorem 9.1.3, we saw that if u is bounded also, then $\alpha = 0$. Does the converse hold?

5. If $f = u + \imath v$ is holomorphic, show that ${}^*du = dv$. In particular, show that ${}^*d(\ln|z-a|) = d(\arg(z-a))$.

6. Let γ be a null homologous cycle in a domain Ω. For any two harmonic functions u_1, u_2 in Ω, show that $\int_\gamma (u_1 {}^*du_2 - u_2 {}^*du_1) = 0$. In particular, show that $\int_\gamma^* du_1 = 0$.

7. Write down full details of the proof of the claim in remark 9.1.2.

9.2 Harnack's Principle:

Taking limits of sequences of known functions is a well established procedure for producing new functions. In this section, we shall use this to produce harmonic functions. This is going to help us in the Dirichlet's problem also.

Lemma 9.2.1 Harnack's inequality : Let u be a non negative real valued function harmonic on Ω and continuous on its boundary. Let $w \in \Omega$ and $\rho > 0$ be such that $B_\rho(w) \subset \Omega$. Then for any $z \in B_\rho(w)$ with $|z-w| = r,\ 0 < r < \rho$, we have,

$$\frac{\rho - r}{\rho + r} u(w) \ \leq \ u(z) \ \leq \ \frac{\rho + r}{\rho - r} u(w). \tag{9.5}$$

Proof: Replacing $u(z)$ by $u(w+z)$, we may as well assume that $w = 0$. The key to the proof of this inequality is to use Poissson Integral Formula (4.50). Clearly, 9.5 is equivalent to

$$\frac{\rho - r}{\rho + r} \int_0^{2\pi} u(\theta) d\theta \leq \int_0^{2\pi} \frac{\rho^2 - r^2}{|\rho e^{\imath\theta} - z|^2} u(\theta) d\theta \leq \frac{\rho + r}{\rho - r} \int_0^{2\pi} u(\theta) d\theta. \tag{9.6}$$

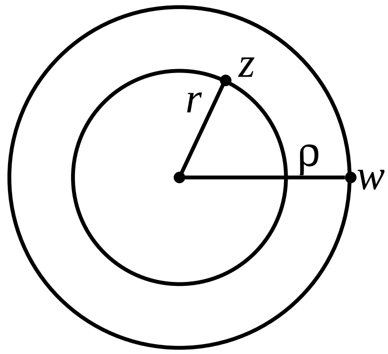


Fig.42

Since, u is a non negative function, this inequality will follow from

$$\frac{\rho - r}{\rho + r} \leq \frac{\rho^2 - r^2}{|\rho e^{\imath\theta} - z|^2} \leq \frac{\rho + r}{\rho - r} \quad \text{for all } \theta. \tag{9.7}$$

But this inequality in turn follows from the inequality

$$\rho + r \geq |\rho e^{\imath\theta} - z| \geq \rho - r. \tag{9.8}$$

This last inequality is obviously true. This completes the proof of the lemma. ♠

Theorem 9.2.1 Harnack's Principle : *Let Ω be the increasing union of a nested sequence of domains $\cdots \subset \Omega_n \subseteq \Omega_{n+1} \subset \cdots$ and let u_n be harmonic in Ω_n such that $u_n \leq u_{n+1}$ on Ω_n. Then u_n converges uniformly on compact sets to either the constant function $+\infty$ or to a harmonic function u on Ω.*

Proof: Let A be the set of all $z \in \Omega$ such that $u_n(z)$ tends to ∞. We shall show that A is both open and closed in Ω. Since Ω is a domain (connected) it then follows that either $A = \Omega$ or $A = \emptyset$.

In either case the compact convergence of the sequence follows from monotonicity and local uniform convergence that we shall see on the way. In the latter case, the harmonicity of the limit function u will follow from the fact that we can represent each u_n by Poisson formula and then take the limit to see that u also has Poisson formula representation.

Let $w \in \Omega$ be any point. Then $w \in \Omega_m$ for some m. Let $\rho > 0$ be such that $B_\rho(w) \subset \Omega_n$ for $n \geq m$. From inequality of (9.5) applied to the harmonic non negative function $u_n - u_m$, we get,

$$\frac{\rho - r}{\rho + r}[u_n(w) - u_m(w)] \leq u_n(z) - u_m(z) \leq \frac{\rho + r}{\rho - r}[u_n(w) - u_m(w)], \tag{9.9}$$

for all z such that $|z - w| = r < \rho$. Suppose now that $w \in A$. Then from the left-hand inequality above, it follows that $u_n(z) \longrightarrow \infty$. Hence $B_\rho(w) \subset A$. This proves that A is open. On the other hand, let $w \notin A$. Then the right-hand inequality above yields that, $z \notin A$. Hence A is closed also.

Thus either $A = \Omega$ or $A = \emptyset$. In the former case, again the left-hand inequality above yields the uniform convergence of u_n to the constant function ∞ on the discs $B_r(w)$. Likewise, in the later case we obtain that u_n converges to u uniformly on $B_r(w)$. This yields compact convergence in either case. ♠

Example 9.2.1 As an illustration of Harnack's principle, and also for future use in the Dirichlet's problem, we shall now construct a harmonic function h on the right-half plane, $\mathbb{G} = \mathbb{G}_0 = \{z : \Re z > 0\}$ with some very specific properties. Let $[\alpha_k, \beta_k]$ be a finite or infinite sequence of disjoint intervals such that $A = \cup_k [\imath\alpha_k, \imath\beta_k]$ is contained in a finite interval $[\imath a, \imath b]$. We are looking for an harmonic function h on $\mathbb{G}$ with the properties:

(i) $h(w) \longrightarrow 0$ as $\Re w \longrightarrow +\infty$;

(ii) $0 \leq h(w) \leq 1$;

(iii) $h(w) \longrightarrow 1$ as w tends to a point on A.

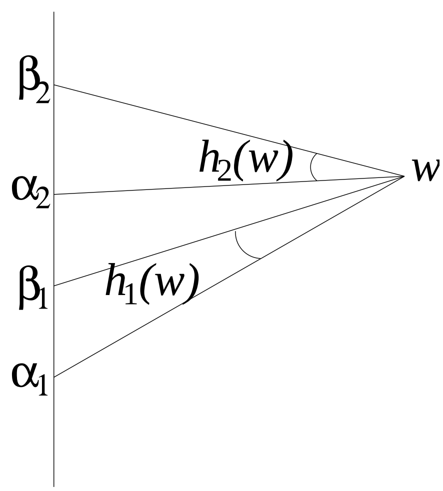


Fig. 43

For any $w \in \mathbb{G}$ consider,

$$h_k(w) = \Im\left(\ln\left[\frac{w - \imath\beta_k}{w - \imath\alpha_k}\right]\right).$$

Then h_k is a harmonic function on $\mathbb{G}$. Geometrically $h_k(w)$ can be thought of as the angle subtended by the segment $[\imath\alpha_k, \imath\beta_k]$ at the point w. Therefore it is easily seen that as $\Re(w) \longrightarrow +\infty$, $h_k(w) \longrightarrow 0$. (This can also be proved rigorously by representing h_k as the definite integral

$$h_k(x + \imath y) = \int_{\alpha_k}^{\beta_k} \frac{x}{x^2 + (y-t)^2} dt.$$

(This we leave to the reader; hint: put $\tan\theta = (t-y)/x$.) Also observe that h_k is positive for all $w \in \mathbb{G}$.

Note that $a < min_k\{\alpha_k\}$ and $b > max_k\{\beta_k\}$. If θ is the angle subtended by the segment $[\imath a, \imath b]$ at w then since the segments $[\imath\alpha_k, \imath\beta_k]$ are disjoint, it follows that $\sum_k h_k(w) < \theta < \pi$. By Harnack's theorem, it follows that $\sum_k h_k$ is a harmonic function in $\mathbb{G}$. We take $h(w) = \sum_k h_k/\pi$. It remains to verify the property (iii) above for h. In fact, if $w \longrightarrow w' \in (\imath\alpha_k, \imath\beta_k)$, then we see that $h_k(w) \longrightarrow \pi$, and $h_j(w) \longrightarrow 0$ for all $j \neq k$. This is easy to see geometrically (but can also be proved rigorously using the above cited integral representation, which we leave to the reader.) This completes the construction of h as required.

9.3 Subharmonic Functions

We have observed, amongst other properties, that the *mean value property* and the *maximum principle* play very important roles in the theory of harmonic functions. This suggests that we should perhaps study the class of functions that possess these properties directly. The first advantage of this approach would be in relaxing the differentiability conditions in the definition of harmonic functions. We shall however keep the continuity hypothesis in tact. In the literature, one finds the theory of harmonic and subharmonic functions which are only *upper semi-continuous.*

Definition 9.3.1 Let u be a continuous real valued function in a domain Ω.
(a) We say u satisfies maximum principle (MP) if in any open subdomain $\Omega' \subseteq \Omega$ the restricted function $u|_{\Omega'}$ does not have a maximum unless it is a constant in Ω'.
(b) We say u has *mean value property (MVP)*, if for every disc $B_r(a) \subset \Omega$, we have,

$$u(a) = \frac{1}{2\pi}\int_0^{2\pi} u(a + re^{\imath t})\, dt. \tag{9.10}$$

Remark 9.3.1

(i) Observe that the condition in (a) is somewhat stronger than saying that the function

u is either a constant or does not have a maximum in the interior of Ω, though often in application we may get away with this much condition only in place of (MP).
(ii) Note that (MVP) is an additive property whereas (MP) is not.

Theorem 9.3.1 *Let Ω be a domain in $\mathbb{C}$, and let u be a continuous real valued function on Ω. Consider the following statements:*
(i) *u satisfies (MP);*
(ii) *u has (MVP);*
(iii) *u is harmonic.*
We have, (i) $\Longleftarrow$ (ii) $\Longleftrightarrow$ (iii).

Proof: (ii) $\Longrightarrow$ (i): Suppose there is $a \in \Omega' \subseteq \Omega$ such that $u(a)$ is the maximum of u on Ω'. If $A = \{z \in \Omega' \ : \ u(z) = u(a)\}$ then clearly A is a closed subset of Ω'. Now given $b \in A$ choose $r > 0$ such that $B_r(b) \subset \Omega'$. Then $u(b) \geq u(b+se^{\imath t}),\ \ 0 \leq t \leq 2\pi,\ \ 0 < s \leq r$ and hence

$$u(b) \geq \frac{1}{2\pi}\int_0^{2\pi} u(b + se^{\imath t})\, dt = u(b). \tag{9.11}$$

It follows that $u(z) = u(b) = u(a)$ on $B_r(b)$. Hence A is open in Ω'. Since Ω' is connected, A must be the whole of Ω', i.e., u is a constant.

The implication (iii) $\Longrightarrow$ (ii) has been proved already.

To prove (ii) $\Longrightarrow$ (iii), let $w \in \Omega$ and let r be chosen so that $B_r(w) \subset \Omega$. Let P_u be the Poisson integral of u restricted to the boundary of $B_r(w)$. It is enough to show that $P_u(w) = u(w)$ on the disc $B_r(w)$.

P_u is harmonic and hence satisfies condition (ii) on its own. It follows that the difference $P_u - u$ has MVP. Hence, as observed above, $P_u - u$ satisfies MP. But P_u is equal to u on $\partial B_r(w)$. Hence, $P_u - u \leq 0$, on $B_r(w)$. Hence $P_u \leq u$. Similar consideration with $u - P_u$ leads us to the conclusion that $P_u \geq u$ on $B_r(w)$. Hence $P_u = u$ in $B_r(w)$. ♠

Remark 9.3.2 One can also consider the so called *minimum principle.* It turns out that this is nothing but the maximum principle for the negative of the function. Also, it should be remarked that a continuous function u which satisfies the maximum and minimum principle need not be harmonic. The simplest example is perhaps, $u(x, y) = x^3$. [This is one reason why we had to work a bit harder in the proof of the implication $(ii) \Longrightarrow (iii)$.] Of course, that is no reason to ignore the class of functions that satisfy the maximum principle. Again, observe that, in the proof of the part $(ii) \Longrightarrow (iii)$ above, we could not have concluded that $P_u - u$ has MP directly from the fact that P_u and u have MP. So, there is a need to strengthen the MP, so that it is 'preserved' under summation.

Before proceeding further, let us take another look backwards. The Laplace's equation in the 1-variable case is nothing but $d^2u/dx^2 = 0$. That is the same as saying that u is a linear function. Now recall that a *convex function* u is one whose graph lies below the line joining $(a, u(a))$ and $(b, u(b))$ for any $a < b$. This leads us, in the two variable case, to consider functions u of the type which take lower values than the harmonic function determined by the boundary value of u, in any domain, i.e, the difference satisfies the maximum principle. But there is a snag: *this apparently begs for the solution of the Dirichlet's problem before hand.* So, there is some need for circumspection.

All these considerations lead us to adopt the following definition, which is somewhat stronger than the maximum principle:

Definition 9.3.2 Let u be a continuous real valued function on a domain Ω. We say u is *subharmonic* if for every subdomain $\Omega' \subseteq \Omega$ and every harmonic function U on Ω' the sum $u + U$ satisfies the maximum principle on Ω'.

Remark 9.3.3

1. Like the definition of continuity, subharmonicity has the local character: *Any function which is subharmonic locally is subharmonic globally.*(Verify this.)

2. A harmonic function is of course subharmonic. Positive multiples of subharmonic functions are subharmonic. Sums of subharmonic function are also subharmonic, but differences need not. And that is why they need to be handled more carefully than harmonic functions.

3. If u has second order partial derivatives then $\bigtriangledown^2 u \geq 0 \Longrightarrow u$ is subharmonic. The converse is not true, in general. However, if the second order partial derivatives are also continuous, then subharmonicity of u implies that $\bigtriangledown^2 u \geq 0$. Take the proofs of all these assertions as exercises.

4. By putting $U \equiv 0$ in the definition, it follows that any subharmonic function satisfies (MP). We have:

Lemma 9.3.1 Let u be a subharmonic function in a domain Ω. Then on every disc Ω such that $\bar{\Omega} \subset \Omega$, we have $u \leq P_u$ where P_u denotes the Poisson function associated to $u|_{\partial\Omega}$.

Proof: The function $(u-P_u)|_\Omega$ by the subharmonicity of u satisfies maximum principle. Hence the maximum of $u-P_u$ has to occur on $\partial\Omega$. By Schwarz's theorem 4.9.7, we know that $u = P_u$ on $\partial\Omega$. Therefore, it follows that $u(z) - P_u(z) \le 0$ for all $z \in \Omega$. ♠

The following theorem gives a characterization of subharmonic functions:

Theorem 9.3.2 *A continuous function u is subharmonic in a domain Ω iff it satisfies the mean value inequality,*

$$u(w) \le \frac{1}{2\pi}\int_0^{2\pi} u(w + r\exp\imath\theta)d\theta, \tag{9.12}$$

whenever $\overline{B_r(w)} \subset \Omega$.

Proof: Note that, any function that satisfies the said inequality (9.12), should satisfy maximum principle. This follows from the proof of (ii) $\Longrightarrow$ (i) in the previous theorem, since we only use the condition (9.12) in (9.11) instead of (9.3.1). Moreover, if u satisfies (9.12), then for any harmonic function U, the function $u+U$ also has this property and hence should have maximum property. This proves the sufficiency of the condition.

The necessity follows from the previous lemma. ♠

The following result tells you that given a subharmonic function on a domain Ω, we can modify it on a small disc so that the new function is now harmonic on the disc retaining of course the subharmonicity on the whole of Ω. This result will be useful later.

Theorem 9.3.3 *Let u be a subharmonic function on Ω, $B_r(w) \subset \Omega$ and let $\tilde{u} = P_u$ on $B_r(w)$ and $= u$ outside of $B_r(w)$. Then $\tilde{u}$ is also subharmonic.*

Proof: By the local character of subharmonicity, it is clear that we have to check the subharmonicity of $\tilde{u}$ only on the boundary points of the disc $B_r(w)$. So, if $a \in \partial B_r(w)$ is any point and $B_\rho(a) \subset \Omega$, let C_1 and C_2 be the portions of $\partial B_\rho(a)$ that lie inside and outside of $B_r(w)$. Then for $z \in C_1$, $\tilde{u}(z) \ge u(z)$ and for $z \in C_2$, $\tilde{u} = u(z)$. Hence

$$\frac{1}{2\pi}\int_0^{2\pi} \tilde{u}(a + \rho\exp\imath\theta)d\theta \ge \frac{1}{2\pi}\int_0^{2\pi} u(a+\rho\exp\imath\theta)d\theta \ge u(a) = \tilde{u}(a).$$

This proves the subharmonicity of $\tilde{u}$. The inequality $u(z) \le \tilde{u}(z)$, $z \in \Omega$ follows from lemma 9.3.1. ♠

Before winding up this section, let us have a slightly stronger version of the maximum principle for future use. For $f : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$, we define

$$\limsup_{x\to x_0} f(x) := \lim_{r\to 0}(\mathrm{Sup}\,\{\mathrm{f(x)\ :\ x \in B_r(x_0) \cap X}\}) \tag{9.13}$$

. By replacing 'Sup' by 'Inf' in the above definition, we get the definition of liminf. It is straight forward to check that $\lim_{x \to x_0} f(x)$ exists iff both $\limsup_{x \to x_0} f(x)$ and $\liminf_{x \to x_0} f(x)$ exist and are equal.

Theorem 9.3.4 *Let u be a continuous function satisfying the maximum principle in a bounded domain Ω. If $\limsup_{z \longrightarrow \zeta} u(z) \leq 0$ for all points ζ in the boundary of Ω, then either $u(z) = 0$ for all $z \in \Omega$ or $u(z) < 0$ for all $z \in \Omega$.*

Proof: Suppose there is $a \in \Omega$ such that $u(a) > 0$. Put $\epsilon = u(a)/2$. For each point $\zeta \in \partial\Omega$, there exists $r > 0$ such that $u(z) < \epsilon$ for $z \in B_r(\zeta) \cap \Omega$. Therefore we get an open set U in $\bar{\Omega}$ such that $\partial\Omega \subset U$ and $u(z) < \epsilon$ for all $z \in U$. Let now K be the set of all points $z \in \Omega$ such that $u(z) \geq u(a)/2$. Then K is a closed subset $\bar{\Omega} \setminus U$ and hence compact. Therefore u attains its maximum on K. But then this is also its maximum on Ω which contradicts the hypothesis. Therefore $u(z) \leq 0$ for all $z \in \Omega$.

Finally, if $u(z_0) = 0$ for some point then it would be its maximum, again violating the hypothesis. Therefore, $u(z) < 0$ for all $z \in \Omega$. ♠

Remark 9.3.4

1. If u_1 and u_2 are subharmonic then so is $u = max\{u_1, u_2\}$.

2. One can define the so called *superharmonic* functions, exactly in a similar manner, by considering minimum principle. However, they are nothing but the negative of subharmonic functions. *Superharmonicity* should not be confused for some property that is stronger than harmonicity. All results formulated about subharmonic functions have analogues for superharmonic functions, which we obtain by merely interchanging the sign and sides of the inequality etc.. For instance, maximum property now becomes minimum property and so on. We shall not even bother to state these results separately and take it as proved the moment the corresponding result is proved for subharmonic functions. However, using theorems 9.3.1 and 9.3.2, it can be seen that a function is harmonic iff it is subharmonic as well as superharmonic.

3. The boundedness of Ω in theorem 9.3.4 is not a necessity. In general, one can work inside the extended plane $\widehat{\mathbb{C}}$ and then the boundary of Ω should be taken in $\widehat{\mathbb{C}}$.

Exercise 9.3

1. Give an example to show that sum of two functions which satisfy MP need not satisfy MP.

2. Prove all the claims made in remark 9.3.3.3.

9.4 Perron's Solution

In this section, we present a solution to the Dirichlet's problem due to O. Perron.[1] Amongst many approaches available in the literature, this one seems to be most general as well as elementary. Recall that given a bounded domain Ω in $\mathbb{C}$ and a continuous function u on its boundary, we want to find a harmonic function in Ω that has limiting values given by u on the boundary of Ω. We first observe that, the problem does not have any solution in the most general form. For example, take $\Omega = \mathbb{D} \setminus \{0\}$, and u to be identically zero on the unit circle and $u(0) = 1$. If there is a harmonic function u on Ω with the desired properties, it then follows that 0 is a removable singularity of u and hence is harmonic in the whole of $\mathbb{D}$. That would contradict the maximum principle. Thus it is necessary to assume some restriction on the nature of the boundary of Ω.

The problem has two parts. Given a domain Ω and a function u on its boundary, we first construct a *candidate* harmonic function on Ω. In the second part, we determine conditions under which this candidate function agrees with the given function at points on the boundary. Obviously, such a condition is '*local*' in nature and has to do with the *geometry* of the domain near this boundary point. Nevertheless, the final answer is breath-takingly beautiful! The only way, Dirichlet's problem may fail to have a solution on a given domain is that the domain has isolated boundary points.

So, let u to be any bounded function on the boundary of a domain Ω. To facilitate notational convenience, we shall denote points of $\partial\Omega$ always by ζ, ζ_0, etc.. Let $u(\zeta) \leq M,\ \forall\, \xi \in \partial\Omega$. Let $\mathcal{B}(u)$ be the set of all subharmonic functions v in Ω such that

$$\limsup_{z\to\zeta} v(z) \ \leq\ u(\zeta),\ \forall\, \zeta \in \partial\Omega. \tag{9.14}$$

This means that given $\epsilon > 0$, there exists $r > 0$ such that $v(z) < u(\zeta) + \epsilon$, in $\Omega \cap B_r(\zeta)$. We set

$$\mathcal{P}_u(z) := \operatorname{lub}\{v(z) \ :\ v \in \mathcal{B}(u)\}, \quad z \in \Omega.$$

[1](1880-1975) Born in Germany, worked in several areas of mathematics such as analysis, number theory, geometry, differential equations etc. He has published more than 200 research articles and wrote several books appreciated by both students as well as teachers. He is well-known for his integrals.

The function $\mathcal{P}_u$ is called the *Perron function* associated to u. A priory, we do not even know whether it is continuous. What we are aiming at is to show that the Perron function is the solution to the Dirichlet's problem.

Theorem 9.4.1 *Let u be a bounded function on the boundary of a domain Ω. Then the Perron function $\mathcal{P}_u$ associated to u is harmonic on Ω.*

Proof: Observe that if $v \in \mathcal{B}(u)$ then, by theorem 9.3.4, it follows that $v(z) \leq M$, where M is a bound for u on $\partial\Omega$. Let B be any disc such that $\overline{B} \subset \Omega$. We want to prove that $\mathcal{P}_u$ is harmonic in B. Let $z_0 \in B$. Choose a sequence $v_n \in \mathcal{B}(u)$, such that $v_n(z_0) \longrightarrow \mathcal{P}_u(z_0)$. Put, $V_n = max\{v_1, \ldots, v_n\}$ for all n. Then by remark 9.3.4 (i) $\{V_n\}$ is a monotonic increasing sequence in $\mathcal{B}(u)$. Form the functions $\widetilde{V_n}$ as in theorem 9.3.3, using the Poisson integral on B. Then we know that $\{\widetilde{V_n}\}$ is also a sequence in $\mathcal{B}(u)$ and by theorem 9.3.3, $v_n(z_0) \leq V_n(z_0) \leq \widetilde{V_n}(z_0) \leq \mathcal{P}_u(z_0)$. Hence $\lim_n \widetilde{V_n}(z_0) = \mathcal{P}_u(z_0)$. On the other hand, by Harnack's principle, $\widetilde{V_n}$ converges to a harmonic limit V on B and clearly $V(z) \leq U(z)$ for $z \in B$ and $V(z_0) = U(z_0)$.

Of course this is not enough to conclude that U is harmonic at z_0. For this, we shall show that $U = V$ on B. Let z_1 be any point in B. As before choose a sequence $w_n \in \mathcal{B}(u)$ such that $w_n(z_1) \longrightarrow U(z_1)$. Now put $W_n = max\{v_1, w_1, \ldots v_n, w_n\}$ and proceed as before to construct $\widetilde{W_n}$ and then pass to the limit W. The result is that we have a harmonic function W in B such that $V \leq W \leq U$ on B and $W(z_1) = U(z_1)$. Clearly, we also have, $V(z_0) \leq W(z_0) \leq U(z_0) = V(z_0)$, and hence the function $V - W$ has a maximum value $= 0$ at z_0. This means that $V = W$ on B and hence, in particular, $U(z_1) = V(z_1)$. This shows that $U = V$ on B. This proves that U is harmonic on Ω, as required. ♠

Definition 9.4.1 Let Ω be a bounded domain and $\zeta_0 \in \partial\Omega$. Let I be a set of positive real numbers such that 0 is a limit point of I. A family $\{\psi_r\}_{r\in I}$ of subharmonic functions taking non positive values is said to be a *family of barriers* at ζ_0 for Ω if given $\delta > 0$, there exist $0 < r < \delta$, and $r \in I$ such that ψ_r is bounded away from 0 on $\Omega \setminus B_r(\zeta_0)$ and $\lim_{z\longrightarrow\zeta_0} \psi_r(z) = 0$. If such a family exists, we say Ω has barriers at ζ_0.

Theorem 9.4.2 *Let Ω be a domain, u be a bounded function on $\partial\Omega$. Suppose that u is continuous at $\zeta_0 \in \partial\Omega$ and that Ω has barriers at ζ_0. Then the Perron's function $\mathcal{P}_u$ associated to u has the property*

$$\lim_{z\to\zeta_0} \mathcal{P}_u(z) = u(\zeta_0).$$

Proof: We have to prove that for every $\epsilon > 0$,

$$\limsup_{z\to\zeta_0} \mathcal{P}_u(z) \leq u(\zeta_0) + \epsilon,$$

and

$$\liminf_{z\to\zeta_0} \mathcal{P}_u(z) \geq u(\zeta_0) - \epsilon.$$

So, let $r > 0$ be such that $\zeta \in B_r(\zeta_0) \Longrightarrow u(\zeta_0) - \epsilon < u(\zeta) < u(\zeta_0) + \epsilon$. (This is where the continuity of u is used.) By taking r smaller we may assume that $\psi := \psi_r$ is a subharmonic function belonging to a family of barriers at ζ_0. Let $m < 0$ be such that $\psi(z) < m$ on $\Omega \setminus B_r(\zeta_0)$. Let M be such that $|u(\zeta)| < M$ for all $\zeta \in \partial\Omega$.

Consider the function

$$V(z) = u(\zeta_0) - \epsilon - \frac{\psi(z)}{m}(M + u(\zeta_0)).$$

This function is clearly subharmonic and we claim that it belongs to $\mathcal{B}(u)$. For if $\zeta \in B_r(\zeta_0) \cap \partial\Omega$, then $V(\zeta) \leq u(\zeta_0) - \epsilon < u(\zeta)$, since ψ is a non positive function. If $\zeta \notin B_r(\zeta_0)$, then $\psi(\zeta)/m \geq 1$ and hence, $V(\zeta) \leq u(\zeta_0) - \epsilon - (M + u(\zeta_0)) = -M - \epsilon < u(\zeta)$. This shows that $V \in \mathcal{B}(u)$ and hence, $\liminf_{z\to\zeta_0} U(z) \geq lim_{z\to\zeta_0} V(z) = u(\zeta_0) - \epsilon$.

To show that $\limsup_{z\longrightarrow\zeta_0} U(z) \leq u(\zeta_0) + \epsilon$, we form

$$W(z) = u(\zeta_0) + \epsilon + \frac{\psi(z)}{m}(M - u(\zeta_0)).$$

Then $-W$ is subharmonic. We shall claim that $\liminf_{z\longrightarrow\zeta} W(z) \geq u(\zeta)$, for all $\zeta \in \partial\Omega$. It follows that for all $v \in \mathcal{B}(u)$, $v - W$ is subharmonic and hence from theorem 9.3.4 we have, $v \leq W$ on Ω. Therefore $\mathcal{P}_u \leq W$ on Ω. In particular this implies that, $\limsup_{z\longrightarrow\zeta_0} \mathcal{P}_u(z) \leq \lim_{z\longrightarrow\zeta_0} W(z) = u(\zeta_0) + \epsilon$, as required.

So, let $\zeta \in B_r(\zeta_0)$. Then for $z \in B_r(\zeta_0)$, $W(z) \geq u(\zeta_0) + \epsilon \geq u(\zeta)$. Therefore $\liminf_{z\longrightarrow\zeta} W(z) \geq u(\zeta)$, in this case. On the other hand, if $\zeta \notin B_r(\zeta_0)$, then for all $z \in \Omega \setminus B_r(\zeta_0)$, we have, $W(z) \geq u(\zeta_0) + \epsilon + M - u(\zeta_0) > u(\zeta)$, and hence again we have, $\liminf_{z\longrightarrow\zeta} W(z) \geq u(\zeta)$. This completes the proof of the claim above and hence that of the theorem also. ♠

Definition 9.4.2 We say a domain Ω is a *Dirichlet's domain* if follwong Dirichlet's problem can be solved (always) on Ω : *Given a continuous function u on $\partial\Omega$, u can be extended to the whole of $\bar{\Omega}$ continuously so that u is harmonic in Ω.*

The first part of the following theorem is an immediate corollary to the above theorem 9.4.2.

Theorem 9.4.3 *A domain Ω is Dirichlet's domain iff it has barriers at each of its boundary points.*

Proof: Inview of theorems 9.4.1 and 9.4.2, we need to show the *only if* part. Suppose Ω is a Dirichlet's domain. Consider,

$$u(\zeta) = \frac{|\zeta - \zeta_0|}{1 + |\zeta - \zeta_0|}.$$

Then u is continuous on $\partial\Omega$. Let f be any (unique) harmonic function on Ω and equal to u on $\partial\Omega$. Then $f(\zeta_0) = 0$. Since $f = u$ is non negative on the boundary, by the minimum principle f is non negative on the whole of $\overline{\Omega}$. In fact $f = u$ is strictly positive on $\partial\Omega$ except at ζ_0, and hence the same is true for f on $\overline{\Omega}$. Hence for any $r > 0$, f is bounded away from 0 on $\Omega \setminus B_r(\zeta_0)$. Therefore, we can take $\psi_r = f$ for all r, as barriers at ζ_0. ♠

Remark 9.4.1 The existence of barriers is easily met in the case of a wide class of geometric situations, for instance, if the boundary of Ω consists of a finitely many piece-wise differentiable smooth arcs, with corners which are not 'too bad'. Thus we can say that the Dirichlet's problem has been solved quite satisfactorily. Below we give two easy instances of this from which it is possible to build-up the existence of barriers in quite a complicated situation also.

Lemma 9.4.1 Let Ω be a region completely contained in the upper half plane with ζ_0 being an isolated point of $\partial\Omega \cap \mathbb{R}$. Then there is a barrier ω at ζ_0.

Proof: In fact take $\omega_r(z) = -\Im z$ for all sufficiently small $r > 0$. ♠

Theorem 9.4.4 *Let Ω be a domain such that each of its boundary point is the end point of a line segment which lies completely outside Ω.* [2] *Then Ω is a Dirichlet's domain.*

Proof: All that we have to prove is the existence of barriers at every point ζ_0 of the boundary. Let A be such a line segment at ζ_0 with its other end point ζ_1. Consider the Möbius transformation

$$T(z) = \frac{z - \zeta_0}{z - \zeta_1}.$$

This maps ζ_0 to 0, ζ_1 to ∞, and the line segment $[\zeta_0, \zeta_1]$ onto the negative part of the real axis. Therefore, T defines a conformal mapping of Ω into a domain contained in $\mathbb{C} \setminus \{r : r < 0\}$. Now for all r, take $\psi_r(z) = -\Re(\sqrt{T(z)})$ and verify that $\{\psi_r\}$ is a family of barriers for Ω at ζ_0. ♠

[2] Such a boundary point is called *linearly accessible*.

Remark 9.4.2 The experience that we gained in the proof of the above theorem tells us that, fractional linear transformations can be used to simplify the geometry near a boundary point of a domain, under certain circumstances. In the following theorem, this idea is exploited to get a complete answer to the existence of barriers and thereby completing the Dirichlet's problem. We shall now consider domains inside the extended plane $\widehat{\mathbb{C}}$. Also, the boundary of such a domain Ω will be taken in $\widehat{\mathbb{C}}$ itself.

Theorem 9.4.5 *Let Ω be a domain $\widehat{\mathbb{C}}$. Suppose that $\zeta_0 \in \partial\Omega$ and there exists at least one more point in the component of $\partial\Omega$ containing ζ_0. Then Ω has barriers at ζ_0.*

Proof: There are two parts to the proof. In the first part, which is rather topological in nature, we shall reduce the problem to a simpler situation. In the second part, this special case is actually proved.

Let C be the component of $\partial\Omega$, that contains ζ_0. Let Ω' be the component of $\widehat{\mathbb{C}} \setminus C$ so that $\Omega \subseteq \Omega.'$ Then Ω' is simply connected, since $\widehat{\mathbb{C}} \setminus \Omega'$ is connected (see theorem 7.4.2). Also, $\zeta_0 \in \partial\Omega'$.

Any finite set in $\widehat{\mathbb{C}}$ is discrete. Since C is connected and has at least two points, C is an infinite set. Let ζ_1, ζ_2 be any two distinct points of C other than ζ_0 and ∞. Let T be the Möbius transformation that takes ζ_0 to $0, \zeta_1$ to ∞ and ∞ to ζ_2. Then $T(\Omega')$ is a simply connected domain in $\mathbb{C}$, containing $T(\Omega)$ and $0 \in \partial(T(\Omega))$. Now, if we construct barriers $\{\psi_r\}$ at 0 for $T(\Omega')$ then they will be barriers for $T(\Omega)$ as well. Since T is a flt, it follows that $\{\psi_r \circ T\}$ will be barriers at ζ_0 for Ω. This is the first part.

Therefore, without loss of generality, assume that Ω itself is simply connected in $\mathbb{C}$, $\zeta_0 = 0 \in \partial\Omega$ and construct a barrier for Ω at 0.

Let now $\ln(z)$ be a well-defined branch of logarithm on Ω. For any $r > 0$, let $f_r(z) = \ln r - \ln(z)$ on $\Omega(0; r) = B_r(0) \cap \Omega$. Then f_r is analytic and $f_r(\Omega(0; r))$ is contained in the right-half plane $\mathbb{G}$. In fact, f_r is continuous on $C_r := \Omega \cap \{w \ : \ |w| = r\}$ and maps it into the imaginary axis. We also observe that as $z \longrightarrow 0$ in Ω we have, $\Re f(z) \longrightarrow +\infty$.

Replace r by a smaller number if necessary and assume that $C_r \neq \emptyset$. It follows that C_r is the disjoint union of a countable number of open arcs. These arcs are mapped by f_r on to mutually disjoint open intervals $(\imath\alpha_k, \imath\beta_k)$ on the imaginary axis. Take $A = \cup_k(\imath\alpha_k, \imath\beta_k)$. Then A is actually contained in an interval of total length less than or equal to 2π. In example 9.2.1, we have a harmonic function h on the right-half plane $\mathbb{G}$ with the properties:

(i) $h(w) \longrightarrow 0$ as $\Re w \longrightarrow +\infty$;

(ii) $0 \leq h(w) \leq 1$;

(iii) $h(w) \longrightarrow 1$ as w tends to a point on A.

Consider $\psi_r(z) = h \circ f_r(z)$ if $z \in \Omega(0;r)$ and $= 1$ otherwise. We claim that $-\psi_r$ is subharmonic in Ω. Clearly ψ is continuous, because of (iii). To show that $-\psi_r$ is subharmonic we appeal to theorem 9.3.2 and verify the mean value inequality. Since inside $\Omega(0,r)$, ψ_r is harmonic and outside this disc it is a constant, it suffices to verify the mean value inequality at $w \in C_r$. If $B_s(w) \subset \Omega$, and L_1 and L_2 are the portions of $\partial B_s(w)$ one contained and the other **not** contained in $B_r(0)$, respectively, then

$$\frac{1}{2\pi}\int_0^{2\pi} \psi_r(w + se^{\imath\theta})d\theta = \frac{1}{2\pi}\left[\int_{L_1} \psi_r(w + se^{\imath\theta})d\theta + \int_{L_2} d\theta\right] \leq 1 = \psi_r(w),$$

from the property (ii). This proves that $-\psi_r$ is subharmonic by theorem 9.3.2. Now it is easily verified that $\{-\psi_r\}$ is a family of barriers at 0. ♠

Exercise 9.4

1. Use the arguments in the above theorem to show that a simply connected domain Ω in $\mathbb{C}$ which is not the whole of $\mathbb{C}$ is a Dirichlet's domain.

2. Let $\Omega = A(0;0,1)$ and let u be the function on $\partial\Omega$ given by $u(\zeta) = 0$ for $|\zeta| = 1$ and $u(0) = 1$. Show that for all $v \in \mathcal{B}(u)$, we have, $v(z) \leq 0,\ \forall\ z \in \Omega$.

9.5 Green's Function

We shall make a little detour to briefly introduce Green's function which its own importance in analysis and in the theory of partial differential equations. At the end, we shall see some connection to Riemann mapping theorem.

Definition 9.5.1 Let D be a domain in $\mathbb{C}$ and $a \in D$. A *Green's function of D with a pole at a* is a function $g_a : D \longrightarrow \mathbb{R}$ such that

(i) g_a is harmonic in $D \setminus \{a\}$;

(ii) $u_a(z) = g_a(z) + \ln|z - a|$ is harmonic in a neighborhood of a;

(iii) $\lim_{z\longrightarrow\zeta} g_a(z) = 0$ for all $\zeta \in \partial D \cup \{\infty\}$.

Remark 9.5.1

(i) Indeed, a Green's function may or maynot exist on a given domain with a prescribed

pole. However, if it exists then it is unique: For, suppose h_a is another functions satisfying (i), (ii) and (iii) above. Then $g_a - h_a$ is harmonic everywhere on D. Also $\lim_{z \longrightarrow \zeta}(g_a - h_a)(z) = 0$ for all boundary points ζ. Therefore by maximum-minimum principle, $g_a - h_a \equiv 0$.
(ii) It follows that if D is a bounded Dirichlet's region, then we can find a harmonic function u_a which coincides with $\ln|\zeta - a|$ on ∂D and take $g_a(z) = u_a(z) - \ln|z-a|$.
(iii) Even though we have said that u_a be harmonic in a neighborhood of a, in the definition above, it follows that u_a is automatically harmonic in the whole of D, because of condition (i) and the fact that $\ln|z-a|$ is harmonic in $\mathbb{C} \setminus \{a\}$.
(iv) Another observation is that g_a is always positive in $D \setminus \{a\}$. This follows easily from the minimum principle using (iii) and the fact that $\lim_{z \longrightarrow a} g_a(z) = +\infty$.
(v) $-\ln|z|$ is the Green's function for $\mathbb{D}$ with its pole at 0.

The following theorem gives an example to illustrate that Green's function need not always exist.

Theorem 9.5.1 *On the domain $\mathbb{C}$ there is no Green's function with pole at any point $a \in \mathbb{C}$.*

Proof: Assume that g_a and u_a are given as above. This means that u_a is a harmonic function on the entire of $\mathbb{C}$. On the other hand, using (ii), we easily see that $u_a(z)/z \longrightarrow 0$ as $z \longrightarrow \infty$. Now from Ex. 6 of 10.2, it follows that u is a constant. But then condition (iii) will be violated. ♠

The following theorem tells us that the Green's function is actually a conformal invariant.

Theorem 9.5.2 *Let D and D' be domains and $f : D \longrightarrow D'$ be a conformal equivalence such that $f(a) = a'$, $a \in D$. If g_a and $g_{a'}$ denote the Green's functions for D and D' with poles at a and a' respectively, then we have $g_{a'} \circ f = g_a$.*

Proof: : We have to merely verify the conditions (i), (ii) and (iii) for the function $g_{a'} \circ f$ on D. The first condition is obvious, since an analytic function followed by a harmonic function produces a harmonic function. Similarly the third condition is also easy because under a conformal mapping f, if $z \longrightarrow \zeta \in \partial D$, then $f(z) \longrightarrow f(\zeta) \in \partial(f(D))$. It remains to show that the function $g_{a'}(f(z)) + \ln|z-a|$ is harmonic in a neighborhood of a. This would follow once we observe that $\ln|f(z)-a'| - \ln|z-a|$ is harmonic in a neighborhood of a. For then, since, $g_{a'}(f(z)) + \ln|f(z) - a'|$ is harmonic in a neighborhood of a', the difference is also so. Finally, recall the fact that the function h defined by

$$h(z) = \begin{cases} \dfrac{f(z) - f(a)}{z - a}, & z \neq a, \\ f'(a) & z = a \end{cases}$$

is analytic. Hence, upon taking the logarithm of the modulus of $h(z)$, we see that $\ln|f(z) - a'| - \ln|z - a|$ is harmonic. ♠

The connection of this result with Riemann mapping theorem has started surfacing now: To find the Green's function for any bounded simply connected region D is the same as finding it for the disc $\mathbb{D}$. This latter problem has been already solved. Can one turn the cart the other way round, i.e., if we know the Green's functions for D, can we determine a conformal mapping of D onto $\mathbb{D}$? This is the question that we take up now, thereby obtaining another proof of Riemann Mapping Theorem.

Exercise 9.5

1. Write down the Green's function g_a for $\mathbb{D}$, $a \in \mathbb{D}$.

2. (i) Let D be a domain bounded by a contour, $a \in D$ and C_1 be a small circle around a contained in D and oriented in the counter clockwise sense. Let $b \in D$ be any point outside this circle. Show that $\displaystyle\int_C (g_a\, {}^\star dg_b - g_b\, {}^\star dg_a) = 2\pi g_b(a)$.
   (ii) Treat the Green's function as a function of two variables, viz., put $g(a, z) = g_a(z)$ so that g is defined on $D \times D \setminus \Delta$ where $\Delta = \{(z, z) \ : \ z \in D\}$ is the diagonal in $D \times D$. Show that g is symmetric.
   (iii) Show that for each fixed $z \in D$, $a \mapsto g(a, z)$ is harmonic in $D \setminus \{z\}$.

9.6 Another Proof of Riemann Mapping Theorem

Let Ω be a simply connected region which is not the whole of the complex plane. As in Step II of the proof of theorem 8.11.1, we can assume that $\Omega \subset \mathbb{D}$.

Let $a \in \Omega$ be any point. We shall first show that there exists a Green's function g_a on Ω with a as the pole. From the exercise 9.4.1 in the previous section, Ω is a Dirichlet's domain. Therefore the continuous function $\ln|\zeta - a|$ on the boundary of Ω, extends to a harmonic function u in Ω. Then $g_a(z) = u(z) - \ln|z - a|$ is a Green's function as required. (Indeed, what we are interested in is the harmonic function u and we could simply have ignore the Green's function associated with it.)

We let ϕ denote an analytic function on Ω with $\Re\phi = u$. (This exists because Ω is simply connected.) We then set $f(z) = (z-a)e^{-\phi(z)}$. Then clearly f is analytic in Ω and $|f(z)| = |z-a|e^{-u(z)}$. Also, it follows that

$$\lim_{z \longrightarrow \zeta} |f(z)| = |\zeta - a| e^{-\ln|\zeta - a|} = 1, \ \zeta \in \partial\Omega. \tag{9.15}$$

A set topological consequence of (9.15) is that f is a proper mapping, i.e., for each compact subset K of $\mathbb{D}$, $f^{-1}(K)$ is compact. (Ex: Prove this.)

Clearly $f(a) = 0$. Indeed, a is the only point mapped onto 0 by f. This fact is going to be crucial in what follows.

The main difficulty is in proving that f is a bijective mapping of Ω onto $\mathbb{D}$. Once we establish that, by multiplying by a suitable constant $e^{\imath\alpha}$, if necessary, we can make $f'(a) > 0$ also. That would complete the proof of the RMT via solution of Dirichlet's problem.

Put $T = \{z \in \Omega \ : \ f'(z) = 0\}$. (Indeed, it turns out that T is an empty set. However, we do not know how to prove this directly without of course using the theorem itself.) Observe that T is a discrete set. Put $\Omega' = \Omega \setminus f^{-1}f(T)$. Then observe that $f|\Omega'$ is also a proper mapping.

The rest of the proof can be completed in two essentially different and equally interesting ways.

Method 1 Here, we begin with the following lemma, which is indeed a result in calculus of two variables. Later, we use the argument principle or to be precise Rouche's theorem.

First we claim that the subset $\Gamma_c := \{z \in \Omega' \ : \ |f(z)| = c\}$ is either empty or a disjoint union of finitely many smooth simple closed curves:
Since f is proper, $f|_{\Omega'}$ is also proper. Therefore, Γ_c is compact or empty. Next, by inverse function theorem, we observe that, at each point z of Γ_c there is a neighborhood U such that $f : U \longrightarrow f(U)$ is biholomorphic. Now $t \mapsto f^{-1}(ce^{\imath t})$ defines a smooth one-one parameterization of $U \cap \Gamma_c$. In particular, this readily implies that each component of Γ_c is simple and having no boundary points. By the compactness of Γ_c finitely many such parametric curves cover Γ_c.

We shall now claim that given $0 < s < 1$, there exists $s < r < 1$ such that $\Gamma_r \cap T = \emptyset$. By the discreteness of T and compactness of Γ_s it follows that $\Gamma_s \cap T$ is a finite set. Again, since Γ_s is compact, we can find an $\epsilon-$ neighborhood V of Γ_s such that $V \cap T \subset \Gamma_s$. Since f is an open mapping there exists $\delta > 0$ such that $W = \{z \ : \ s-\delta < |z| < s+\delta\} \subset f(V)$. By uniform continuity of f on the compact set Γ_s, we can choose $\delta > 0$ sufficiently small such that $f^{-1}(W) \subset V$. Now take r such that $s < r < min\{s + \delta, 1\}$. Then r is as required.

Now, given any $w_0 \in \mathbb{D}$, choose r such that $|w_0| < r < 1$ such that $\Gamma_r \cap T = \emptyset$. It follows that Γ_r is a disjoint union of finitely many simple closed contours. Let γ_r be the cycle obtained by tracing each of the components of Γ_r exactly once in the counter clockwise sense. Then it follows that the winding number $\eta(\gamma_r, z)$ is either 0 or equals 1, for all point of $\Omega \setminus \Gamma_r$, since, the same is true for the circle $|w| = r$ and points of $\mathbb{D}$. We now apply Rouché's theorem. Our aim is to show that the equation $f(z) = w_0$ has precisely one solution inside Ω. Our knowledge says this is the case for $w = 0$. So we take $g(z) = f(z) - w_0$, and observe that $|f(z) - g(z)| = |w_0| \leq |f(z)|$ for all points $z \in \Gamma_r$. Therefore Rouche's theorem yields the desired result. This completes the proof of RMT by the first method.

Method 2: This method is more point-set topological. It may also be called algebraic topological, since it employs the notion of covering spaces.

Using properness, one first observes that f is surjective (exercise 2 below). By Inverse Function Theorem, f is a local homeomorphism on $\Omega \setminus T$. Therefore, it follows that $f : \Omega' \longrightarrow \mathbb{D} \setminus f(T)$ is a proper surjective local homeomorphism. Also since $f^{-1}(f(T))$ is discrete, Ω' is connected. Hence, by covering space theory, $f : \Omega' \longrightarrow \mathbb{D} \setminus f(T)$ is a covering projection. But $f^{-1}(0) = \{a\}$. That means every fibre of f has exactly one point in it, i.e., $f : \Omega' \longrightarrow \mathbb{D} \setminus f(T)$ is a homeomorphism. Now by continuity of f, it is easily seen that $f : \Omega \longrightarrow \mathbb{D}$ is also injective. That completes another proof of RMT. ♠

Exercise 9.6

1. Let $f : \Omega \longrightarrow \mathbb{D}$ be a continuous mapping such that for each $\zeta \in \partial\Omega$, we have $\lim_{z\to\zeta} |f(z)| = 1$. Then show that f is a proper mapping.

2. Let $f : \Omega_1 \longrightarrow \Omega_2$ be a continuous, proper, open mapping of two domains. Show that f is surjective.

3. Write down details of the last step in the proof above, claiming injectivity of f on the whole of Ω.

9.7 Multi-connected Domains

We shall now see how the solution of Dirichlet's problem can be employed in the conformal classification of multi-connected domains. To begin with let us opt a convenient definition.

Definition 9.7.1 We shall say that a domain Ω is of *n-connectivity* or is *n-fold connected* if $\widehat{\mathbb{C}} \setminus \Omega$ has precisely n connected components. For $n = 1$, this corresponds to simply connectedness. If $n \geq 2$ we also use the terminology *multi–connectedness.*

Of course n can be infinity also in the above definition. For the sake of simplicity of the exposition, we shall restrict to finite ($n < \infty$) connectivity only.

In order to employ the solution of Dirichlet's problem, we shall have to put the restriction that no component of $\widehat{\mathbb{C}} \setminus \Omega$ is a singleton. Indeed, this is not a handicap at all. For, if some components are singletons, then we can fill them up, classify conformally the larger domain so obtained, and then delete the images of these singletons from the objects so obtained. The justification in doing this is that if $f : \Omega \longrightarrow \Omega'$ is a biholomorphic map, then f defines a bijection of the singleton boundary components of Ω with those of Ω'. Moreover, f extends to a conformal mapping of the domains obtained by filling up any set of corresponding singleton sets on either side. (Exercise 1. Supply proof of all this.)

The next step is to choose a set of canonical domains which are going to represent the conformal classes. In the case of simply connectedness, we chose the complex plane and the unit disc. We could have chosen the upper half plane in place of the unit disc. Thus, clearly there is no unique choice in general. Indeed, the choices available are much more for $n \geq 2$ and what we choose is just a matter of our taste.

We shall now discuss the case $n = 2$ completely. This may help to grasp the ideas involved while keeping the complexity of the situation to the minimum. To begin with notice that all punctured discs of finite radius $A(w; 0, r)$ are biholomorphic to each other. The domain $\mathbb{C}^*$ however stands alone. (See Exercise 5 in Misc 5.8). In what follows according to our convention above, we shall consider annuli with inner radius positive. (See 5.12 for the notation.)

Theorem 9.7.1 *The annuli $A(a; r_1, r_2)$ and $A(b; R_1, R_2)$ are conformally equivalent iff* $\dfrac{r_1}{r_2} = \dfrac{R_1}{R_2}$.

Proof: The 'if' part was easily seen by taking $z \mapsto \lambda z + b - a$ where $\lambda = R_1/r_2 = R_2/r_2$. To see the 'only if' part we proceed in the following way. First of all, without loss of generality, we may assume that $a = 0 = b$.

Let $\psi : A(0; r_1, r_2) \longrightarrow A(0; R_1, R_2)$ be a biholomorphic mapping. Let C_r denote the circle of radius r, $r_1 < r < r_2$. Since $A(0; r_1, r_2) \setminus C_r$ has two connected components, say, S_1, S_2, it follows that $A(0; R_1, R_2) \setminus \psi(C_r)$ has two components, say T_1, T_2. Let us label them in such a way that the closure of S_1 contains the circle of radius r_1 etc.. Now there are precisely two possibilities:
(a) $\psi(S_1) = T_1$ and $\psi(S_2) = T_2$; or
(b) $\psi(S_1) = T_2$ and $\psi(S_2) = T_1$.
In the latter case, we consider the mapping $z \mapsto 1/\psi(z)$ which is a biholomorphic mapping of $A(0; r_1, r_2)$ to $A(0; 1/R_2; 1/R_1)$. Since we have to show that $r_1/r_2 = R_1/R_2$, we can as well assume that our ψ itself is such that (a) holds.

Now, since ϕ is a proper mapping it follows that $\lim_{|z| \to r_j} |\phi(z)| = R_j, j = 1, 2$. Choose $\alpha, \beta \in \mathbb{R}$ so that

$$\alpha r_j + \beta = R_j, \quad j = 1, 2$$

and consider the harmonic function $u(z) = \alpha \ln |z| + \beta$. Then the harmonic function $v(z) = \ln |\psi(z)| - u(z)$ on the annulus $A(0; r_1, r_2)$ has the property that

$$\lim_{|z| \to r_j} v(z) = 0, \quad j = 1, 2.$$

From the extended maximum principle (theorem 9.3.4), it follows that $v(z) \equiv 0$ and hence

$$\ln |\psi(z)| = \alpha \ln |z| + \beta, \quad z \in A(0, r_1, r_2). \tag{9.16}$$

This means that ψ maps circles C_r onto circles C_R, where $\ln R = \alpha \ln r + \beta$.

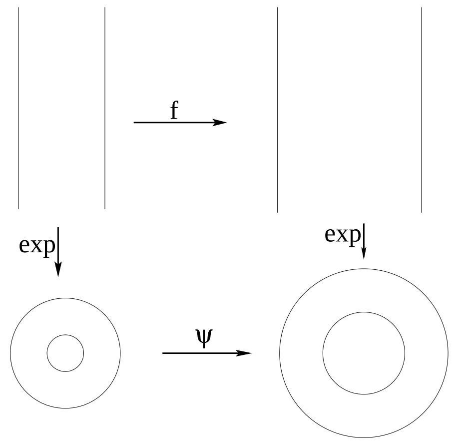


Fig. 44

Now consider the exponential map $\exp : z \mapsto e^z$. This maps the infinite strip

$$S(c_1, c_2) := \{z \ : \ \ln c_1 < \Re z < \ln c_2\}$$

onto the annulus $A(0; c_1, c_2)$. Since $S(c_1, c_2)$ is simply connected, given any holomorphic mapping $\psi : A(0; r_1, r_2) \longrightarrow A(0; R_1, R_2)$, it follows that there is a well defined mapping $f : S(r_1, r_2) \longrightarrow S(R_1, R_2)$ such that $\exp\ f(z) = \psi(\exp z)$. Further, if ψ is biholomorphic, we apply the same to its inverse, ψ^{-1} to get $g : S(R_1, R_2) \longrightarrow S(r_1, r_2)$ such that $\exp\ g(z) = \psi^{-1}(\exp\ z)$. Hence, $\exp\ f(g(z)) = \exp\ z,\ \ z \in S(R_1, R_2)$. This means $f \circ g(z) = z + 2\pi \imath n, z \in S(R_1, R_2)$, for some integer n. By subtracting the constant $2\pi\imath n$ from f, we may as well assume that $f \circ g(z) = z$ for all $z \in S(R_1, R_2)$. In particular, we have proved that f is biholomorphic.

Since $\exp\,(f(z)) = \psi(\exp\ z)$, it follows that given $z \in S(c_1, c_2)$, $f(z + 2\pi\imath) = f(z) + 2m_z\pi\imath$ for some integer m_z. Moreover, the integer m_z is the same when z varies over a 'small' open set. Since $S(r_1, r_2)$ is connected, this means that we have a well defined integer m such that $f(z + 2\pi\imath) = f(z) + 2m\pi\imath$. Therefore $f(z + 2k\pi\imath) = f(z) + 2km\pi\imath$ for integrs k. Similarly, for g, we get an integer n such that $g(z + 2k\pi\imath) = g(z) + 2kn\pi\imath$ or all integers k. Now $f \circ g = Id$ means that $mn = 1$. Therefore, $m = \pm 1$. Since f maps $x = \ln r_j$ to $x = \ln R_j, j = 1, 2$ and preserves orientations, we can further conclude that $m = 1$, which means

$$f(z + 2\pi\imath) = f(z) + 2\pi\imath. \tag{9.17}$$

If we put $f(x, y) = U(x, y) + \imath V(x, y)$, it now follows from (9.16) that U is independent of y. Since U is harmonic, this means $U(x, y) = ax + b$. Since V is a harmonic conjugate, we get $V(x, y) = ay + c$. On the other hand (9.17) implies that $V(x, y + 2\pi\imath) = V(x, y) + 2\pi\imath$.

Therefore $a = 1$. Since $U(\ln r_j, y) = \ln R_j$, it follows that $b = \ln R_1 - \ln r_1$ and $\ln R_2 = \ln r_2 + b$. Therefore $\ln R_2 - \ln R_1 = \ln r_2 - \ln r_1$ which proves the claim $R_2/R_1 = r_2/r_1$. ♠

Thus, amongst all annuli, we need to choose only those of the form $A(0; 1, r)$ for $1 < r \leq \infty$. Observe that we allow $r = \infty$ in the above list. Indeed $A(0; 1, \infty)$ is conformal with the punctured unit disc. In conclusion, the family of all conformally equivalent 2-connected domains can be parameterised by the half open interval $(1, \infty]$ except one member viz. the class of $\mathbb{C}^*$.

In order to complete the classification of 2-fold connected domains, we should show that every such domain is conformal with one of the above listed annuli. The argument that we put forth is the same for the general case also and hence we shall directly deal with domains of n-connectivity and show that every n-connected domain is conformal with a domain obtained from an annulus $A(0; r_1, r_2)$ by removing $(n-3)$ concentric circular arcs. The arguments are broken up into a number of steps. We begin with a tentative definition.

Definition 9.7.2 A domain Ω in $\widehat{\mathbb{C}}$ is said to be *nice* if it is obtained by removing from $\widehat{\mathbb{C}}$, n simply connected domains bounded by mutually disjoint smooth simple closed contours, $n \geq 0$.

It is clear that such an n-fold connected domain is nice then $\widehat{\mathbb{C}} \setminus \Omega$ has n components each of them being simply connected domains. Indeed we have:

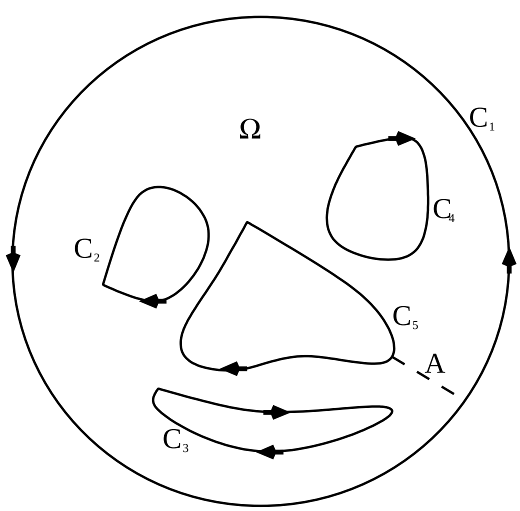


Fig. 45

Lemma 9.7.1 Let Ω be a nice, n-fold connected domain with $C_1, \dots, C_n$ as boundary components oriented in such a way that Ω lies to the left of each of the curves C_j. Then
(i) $\sum_{j=1}^{n} C_j$ is homologous to 0 in $\bar{\Omega}$.
(ii) Every cycle in $\bar{\Omega}$ is homologous to a cycle of the form $\sum_{j=1}^{n} m_j C_j$, with $m_j \in \mathbb{Z}$.

Proof: The argument is similar to that in Exercise 4 of 7.4 and left to the reader (see Fig. 45). ♠

Step 1 In the first step, we shall show that Ω is conformally equivalent to a nice domain.

So, let $C_1, C_2 \cdots, C_n, n \geq 2$ be the boundary components of a domain Ω, none of which is a singleton. Let Ω_1 be the component of $\widehat{\mathbb{C}} \setminus C_1$ which contains Ω. If $\Omega_j, j = 2, \dots,$ are other components of $\widehat{\mathbb{C}} \setminus C_1$, it is clear that $\bar{\Omega}_j \cap C_1 \neq \emptyset$. In particular, it follows that $\bar{\Omega}_j \cup C_1$ is connected. Therefore $\cup_{j \geq 2} \bar{\Omega}_j \cup C_1$ is connected. This set is equal to $\widehat{\mathbb{C}} \setminus \Omega_1$. Therefore, Ω_1 is simply connected. (See theorem 7.4.2). Clearly, it is not the whole of $\mathbb{C}$. By Riemann mapping theorem, there exists a biholomorphic mapping $T_1 : \Omega_1 \longrightarrow \mathbb{D}$. Under this mapping Ω is sent to a domain $T_1(\Omega)$ inside $\mathbb{D}$ with one of its boundary component being the unit circle. Indeed, this one corresponds to the boundary component C_1 of Ω. Observe that, we are not claiming that T_1 defines a homeomorphism of C_1 with the unit circle.) The net result is that we can now replace Ω by $T_1(\Omega)$ and assume that Ω itself had one of its boundary components say, C_1 a smooth simple closed curve. One by one, we repeat the above process with each component C_k. This time under T_k the boundary component C_k will correspond to the unit circle. However, under this mapping all other boundary components C_i are mapped biholomorphically onto some boundary component of $T(\Omega)$. In particular, those which are smooth simple closed curves remain to be smooth simple closed curves. Repeating this process n times, we achieve our claim.

From now onwards, we shall assume that Ω itself is nice. Indeed, in order to have a nice picture, we can further assume that Ω is contained in the unit disc $\mathbb{D}$ and $C_1 = \partial\mathbb{D} = \{|z| = 1\}$. By changing the labeling if necessary, we may assume that C_1 is oriented anti-clockwise and for $2 \leq k \leq n$, C_j are oriented clockwise. (See figure 45.)

Let $\gamma = C_1 + \cdots + C_n$. Then γ bounds Ω and $\eta(\gamma; z) = 1$ iff $z \in \Omega$.

Step 2 In this step, we obtain harmonic functions $\omega_k, 1 \leq k \leq n$ in a neighborhood of $\overline{\Omega}$ such that

$$\omega_k(\zeta) = \begin{cases} 1, & \zeta \in C_k \\ 0, & \zeta \in C_j,\ j \neq k. \end{cases}$$

Observe that Ω is clearly a Dirichlet's domain. Hence, there exist harmonic functions ω_k on Ω as claimed. It remains to see how to extend them to a neighborhood of $\overline{\Omega}$. For this, we again use the fact that, we can find a conformal map ψ_i of Ω onto some domain such that a given boundary component C_k in mapped onto the unit circle. Now, using the reflection principle, we can extend the harmonic functions $\omega_j \circ \psi_j^{-1}$ across the unit circle and then go back to the domain Ω to obtain the extensions of ω_i. Do this one by one to each boundary component.

Clearly, by Minimum-Maximum principle, for all $1 \leq k \leq n$, we have,

$$0 < \omega_k(z) < 1 \;\;\&\;\; \sum_k \omega_k(z) = 1, \quad z \in \Omega.$$

We now put $\alpha_{jk} = \displaystyle\int_{C_j} {}^*d\omega_k$, where ${}^*d\omega_k$ are the conjugate harmonic differentials. (See remarks 9.1.2.) Let $A = ((\alpha_{jk}))$, $1 \leq j \leq n$, $1 \leq k \leq n-1$, be the $n \times (n-1)$ matrix.

Step 3 In the third step, we claim that the following linear system of equations has a solution:

$$A\Lambda = (2\pi, 0, \ldots, 0, -2\pi)^t \tag{9.18}$$

We first observe that the $(n-1) \times (n-1)$ matrix B, obtained by deleting the last row of A is non singular. For suppose $B\Lambda' = 0$, where $\Lambda' = (\lambda_1', \ldots \lambda_{n-1}')^t$. Consider the harmonic function $\omega' = \displaystyle\sum_{k=1}^{n-1} \lambda_k' \omega_k$. Then the conjugate differential ${}^*d\omega'$ has all its fundamental periods vanishing, i.e.,

$$\int_{C_j} {}^*d\omega' = \sum_k \lambda_k' \alpha_{jk} = 0.$$

Therefore, it follows from lemma 9.7.1 that ${}^*d\omega'$ is an exact 1-form. i.e., ω' has a well defined conjugate on $\overline{\Omega}$ and hence is the real part of an analytic function g therein (see remark 9.1.2). Clearly, under g, each contour C_k is mapped onto a vertical line segment since the real part of it is a constant there. We shall show that f must be constant. Let w be any point **not** on any of these vertical line segments. If $f(z) = w$ for some $z \in \Omega$, then since, $\eta(\gamma, z) = 1$, we must have, by the argument principle, that $\eta(f(\gamma), w) \neq 0$. But this is absurd since $f(\gamma)$ is a union of line segments. Therefore, $f(\overline{\Omega})$ is contained in the union of line segments. But $\overline{\Omega}$ is connected and hence $f(\overline{\Omega})$ is contained in a single vertical line segment. Therefore ω' is a constant. This implies that $\lambda_k' = 0$ for all k, as required (we shall leave this as a simple exercise).

Thus we have proved that the matrix B is non singular. It follows that $B\Lambda_0 = (2\pi, 0, \ldots, 0)$ has a unique solution $\Lambda_0 = (\lambda_1, \ldots, \lambda_{n-1})^t$. We take $\omega = \sum_{k=1}^{n-1} \lambda_k \omega_k$. Since $\gamma = \sum_j C_j$ is null homologous in $\bar{\Omega}$, we have, $\displaystyle\sum_{j=1}^{n-1} \int_{C_j} {}^*d\omega = -\int_{C_n} {}^*d\omega$. (To prove this use Exercise 9.1.6.) It follows that $\sum_{j=1}^{n-1} \lambda_j \alpha_{jn} = -2\pi$. Taking λ_n to be any real number say $= 0$, it follows that $\Lambda = (\lambda_1, \ldots, \lambda_{n-1}, \lambda_n)^t$ is a solution of (9.18) as required.

Now the conjugate differential ${}^*d\omega$ of the harmonic function $\omega = \sum \lambda_k \omega_k$ has vanishing periods on all cycles $C_k,\ k \neq 1, n$, and on these two cycles it has periods $\pm 2\pi$. Choose a polygonal arc A in Ω, such that $\bar{A}$ joins a point of C_n with a point of C_1. (See figure 45.) It follows that ${}^*d\omega$ has all its period vanishing in $\Omega \setminus \bar{A}$. Therefore, there is a harmonic function $\tilde{\omega}$ on $\Omega \setminus \bar{A}$ such that $d(\tilde{\omega}) =^* d\omega$. We put $f = \omega + \imath\tilde{\omega}$. Then f is a holomorphic function on $\Omega \setminus \bar{A}$. Also, at any point of $a \in A$ the limit of $f(z)$ as $z \to a$ from the two sides of A differ by an integral multiple of $2\pi\imath$ being the period of ${}^*d\omega$ on the cycle C_1. Therefore, the function $\psi(z) = e^{f(z)}$ is a single valued holomorphic function on $\Omega \setminus A$ and is continuous on A. But then ψ should be holomorphic on A as well (see exercise 4.5.4 and 4.5.5). As before using reflection principle, we may assume that ψ is actually holomorphic on an open set containing $\bar{\Omega}$.

Step 4 In this final step, we claim that ψ is a conformal mapping of Ω into the annulus $1 < |w| < e^{\lambda_1}$, with its image avoiding precisely $n-2$ concentric arcs on the circles $|w| = e^{\lambda_k},\ 2 \leq k \leq n-2$.

Introduce the notation $m_j(w) := \dfrac{1}{2\pi\imath}\displaystyle\int_{C_j} \frac{\psi'(z)dz}{\psi(z) - w}$ for all w not in $\cup_{j=1}^n \psi(C_j)$. Observe that each m_k is defined for $w \notin \psi(C_k)$. Therefore the number of solutions $z \in \Omega$ of the equation $\psi(z) = w$ for $w \notin \cup_{k=1}^n \psi(C_k)$ is given by

$$\frac{1}{2\pi\imath}\int_\gamma \frac{\psi'(z)dz}{\psi(z) - w} = \frac{1}{2\pi\imath}\sum_{k=1}^n \int_{C_k} \frac{\psi'(z)dz}{\psi(z) - w} = m_1(w) + \cdots + m_n(w) \tag{9.19}$$

Observe that, since $arg\,\psi(z)$ has period 2π around C_1, we have,

$$\int_{C_1} \frac{\psi'(z)dz}{\psi(z)} = 2\pi\imath.$$

Likewise, we have,

$$\int_{C_n} \frac{\psi'(z)dz}{\psi(z)} = -2\pi\imath, \text{ and } \int_{C_k} \frac{\psi'(z)dz}{\psi(z)} = 0,\ \forall\ k \neq 1, n.$$

This amounts to saying that in (9.19) if we put $w = 0$, then the first and the last term on RHS are 1 and -1 respectively, and all other terms are 0. Since m_j are locally constants, we have, $m_1(w) = 1$ for all $|w| < e^{\lambda_1}$ and equal to 0 for $|w| > e^{\lambda_1}$. Similarly, $m_n(w) = 0$, for $|w| > 1$ and $= 1$ for $|w| < 1$. On the other hand, take any $k = 2, \ldots, n-1$. Then for $w_0 \notin \psi(C_k)$, $m_k = 0$. (Note that m_k is not defined on $\psi(C_k)$.)

Now suppose $w_0 \in \psi(\Omega) \setminus \cup_{k=1}^n \psi(C_k)$. Then clearly $\sum_k m_k(w_0) > 0$. This is possible only if $m_1 = 0$ and $m_n = 0$ i.e., $\sum_k m_k(w_0) = 1$. Therefore, $1 \leq |w_0| \leq e^{\lambda_1}$. By continuity, it follows that $\psi(\Omega) \subset \overline{A(0,1,e^{\lambda_1})}$. Once again, since $\psi(\Omega)$ is non empty open, it follows that $\psi(\Omega) \subset A(0;1,e^{\lambda_1})$. In particular, this implies $0 < \lambda_1$. Since $\psi(C_k) \subset \{|z| = e^{\lambda_k}\}$, this also means that $0 \leq \lambda_j \leq 1,\ 2 \leq j \leq n$.

We have already proved that for every $w_0 \in A(0;1,e^{\lambda_1}) \setminus \cup_{k=2}^{n-1} \psi(C_k)$, $\sum_k m_k(w_0) = 1$. This is the same as saying

$$\psi : \Omega \to A(0;1,e^{\lambda_1}) \setminus \cup_{k=2}^{n-1} \psi(C_k)$$

is a bijection.

We now claim that $\psi(C_j) \cap \psi(C_k) = \emptyset, j \neq k$. Choose a simple closed curve τ in Ω which 'separates' C_j and C_k. (See Exercise 7.5.17 and 7.5.18.) Then ψ maps the two components of $\Omega \setminus \tau$ onto the two components X, Y on either side of $\psi(\tau)$. By continuity, it follows that $\psi(C_j)$ and $\psi(C_k)$ are in the closure of X and Y respectively, say. However, it is clear that they are disjoint from $\psi(\tau)$. Therefore $\psi(C_j) \cap \psi(C_k) = \emptyset$.

In particular it follows that $0 < \lambda_j < \lambda_1, j = 2, 3, \ldots, n-1$. It now follows that since $\cup_{k=2}^{n-1} \psi(C_k)$ does not separate the annulus $A(0;1,e^{\lambda_1})$. In particular, each $\psi(C_k)$ is an incomplete arc of a circle with center at 0 and all are contained in the interior of the annulus $A(0;1,e^{\lambda_1})$. This completes the step 4 and hence establishes that every n-connected domain is biholomorphic to an annulus with $n-2$ slits.

It remains to determine which of these slit-open annuli are conformal amongst themselves. The first thing we do is to normalize so that all the annuli have center at 0 and their inner circles all have radius 1. For simplicity, recall the case $n = 2$. Here the only freedom now is to choose the radius r_1 of the outer circle. Thus we can say that the 'space' of all conformal classes of 2-fold connected planar domains other than $\mathbb{C}^*$, can be 'parameterized' by the open ray $X_1 := \{r_1 \in \mathbb{R} \ :\ r_1 > 1\}$. Next consider the case $n = 3$. We can perform rotations so that the slit sweeps the angular sector $0 \leq \theta \leq t$, for some $0 < t < 2\pi$. Of course, the radius of the slit could be any thing between 1 and r_1, where r_1 is the radius of the annulus. It remains to see that two such slit open annuli are not biholomorphic to each other. This is left to the reader.

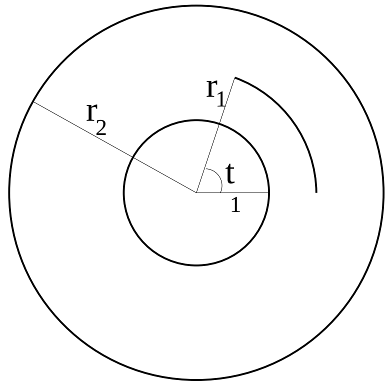


Fig. 46

Thus we see that the space of all conformal classes of 3-fold connected planar regions can be parameterized by the subspace $X_2 := \{r_1, t, r_2) \in X_1 \times (0, 2\pi) \times \mathbb{R} \ : \ 1 < r_2 < r_1\}$. Clearly, this space is of real dimension 3. For $n > 3$, for each of the additional $n-3$ slits, we have the freedom to choose the three real numbers, one determining the radius and the other two determining its angular position. Of course, there are certain 'open' constraints these real numbers are subjected to but it is fairly obvious that the space of all conformal classes of n-fold connected planar region can be parameterized by $3n-6$ real variables.

Exercise 9.7

1. Show that $\alpha_{jk} = \alpha_{kj}$ [Hint: Use the fact that $\int_\gamma (\omega_k \, {}^*dw_j - \omega_j \, {}^*dw_k) = 0$, (Ex. 6 of 9.1).]

2. Write down a complete parameterization of the space of all conformal classes of 4-fold connected planar regions.

9.8 Miscellaneous Exercises to Ch. 9

1. Show that $u(z) = \Im \left[\dfrac{1+z}{1-z}\right]^2$ is harmonic in the open unit disc $\mathbb{D}$. Also show that $\lim\limits_{r \longrightarrow 1^-} u(re^{\imath\theta}) = 0$ for all θ. Why this does not contradict the maximum principle?

2. Let f be an analytic function in an annulus $r_1 < |z| < r_2$, and let $M(r)$ denote the maximum of $|f(z)|$ on $|z| = r$. By considering a suitable linear combination of $\ln|f(z)|$ and $\ln|z|$ show that

$$M(r) \leq M(r_1)^a / M(r_2)^{1-a},$$

where $a = \dfrac{\ln(r_2/r)}{\ln(r_2/r_1)}$. This is known as *Hadamard's three circles theorem.*

3. If $\alpha = 0$ in theorem 9.1.1, does it imply that u is harmonic on the entire disc?

4. Let $g : \overline{\mathbb{D}} \longrightarrow \mathbb{C}$ be a continuous function. Define $g_r : \mathbb{S}^1 \longrightarrow \mathbb{C}$, by $z \mapsto g(rz),\ \ 0 < r < 1$. Then show that g_r tends to $g|_{\mathbb{S}^1}$ uniformly on $\mathbb{S}^1$, as $r \longrightarrow 1^-$.

5. Let $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$ be a continuous function, and $\tilde{f}$ be its extension to the interior of the circle given by

$$\tilde{f}(re^{\imath\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{\imath\theta}) P_r(\theta - t) dt.$$

Then show that $\tilde{f}$ is continuous on $\overline{\mathbb{D}}$ and both its real and imaginary parts are harmonic in $\mathbb{D}$. Define $\tilde{f}_r(z) = \tilde{f}(rz)$, for $|z| = 1$. Then show that for each $0 \leq r < 1$, there exist a sequence of polynomial in z and $\bar{z}$ uniformly converging to $\tilde{f}_r$ on $\mathbb{S}^1$. [Hint: Use exercises 1-3 at the end of section 10.2.]

6. **Weierstrass's Approximation Theorem :**
(a) Let $f : \mathbb{S}^1 \longrightarrow \mathbb{C}$ be a continuous function. Then there exists a sequence of Laurent polynomials $\{p_n(z, z^{-1})\}$ uniformly convergent to f on $\mathbb{S}^1$. If f is real valued then the coefficients of p_n can be taken to be real. [Hint Use ex. 4-5. Compare Example(8.6.1).]
(b) Let $g : [0,1] \longrightarrow \mathbb{C}$ be a continuous function. Then there is a sequence of polynomial functions $\{p_n(t)\}$ converging uniformly on $[0,1]$ to $g(t)$.

7. *Geometric interpretation of Poisson integral.*
(a) Fix $z \in \mathbb{D}$. Consider the flt:$S(w) = \dfrac{w - z}{1 - \bar{z}w}$. Show that if $|w| = 1$, then $|S(w)| = 1$ and $w, z, S(w)$ are collinear.
(b) Show that $(w - z)(\overline{S(w) - z}) = |z|^2 - 1$
(c) Write $w = e^{\imath\theta}$ and $S(w) = e^{\imath\vartheta}$. Differentiate $(\star)$ to obtain,

$$\frac{d\vartheta}{d\theta} = \left|\frac{S(w) - z}{w - z}\right|.$$

Finally show that

$$P_U(z) = \frac{1}{2\pi}\int_0^{2\pi} U(\theta)d\vartheta = \frac{1}{2\pi}\int_0^{2\pi} U(\vartheta)d\theta.$$

8. Consider $Har(\Omega) \subset C(\Omega, \mathbb{R})$, the subset of all continuous functions on the closed disc which are harmonic in the interior. Show that it is a closed subspace under the sup norm, and hence is a complete metric space.

9. Assume that $U(\xi)$ is piecewise continuous and bounded on the real line. Show that

$$Q_U(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{y}{(x-\xi)^2 + y^2} U(\xi)d\xi$$

represents a harmonic function in the upper half plane. Also, show that if U is continuous at some point $\xi_0 \in \mathbb{R}$ then $\lim_{z \longrightarrow \xi_0} Q_U(z) = U(\xi_0)$.

10. Let $u(z)$ be harmonic in $\boldsymbol{H}$ and $|u(z)| \leq Ky$ for all $y > 0$. Then show that there exist k such that $|k| \leq K$, and $u = ky$ for $y > 0$.

11. Show that any automorphism of the annulus $A(0; 1, R)$ is the form

$$z \mapsto R\frac{e^{\imath\theta}}{z};\ \text{OR } z \mapsto e^{\imath\theta}z, \theta \in \mathbb{R}$$

Hence deduce that this group is isomorphic to $\mathbb{S}^1 \times \mathbb{Z}_2$.

Chapter 10

PERIODIC FUNCTIONS

Functions such as exp, sin, cos, tan etc. stand out amongst all holomorphic (meromorphic) functions due to their rich properties some of which may be attributed to the fact that they obey

$$f(z + 2\pi\imath) = f(z), \tag{10.1}$$

in the domain of their definition. Functions satisfying (10.1) are called *periodic functions* with *period* 2π. In this chapter, we shall initiate the study of meromorphic functions with property silimar to (10.1) with a modest aim. We shall be able to cover only a fraction of a vast subject. As an application, we shall then prove the so called 'Big Picard Theorem' 10.9.6.

10.1 Singly Periodic Functions

Definition 10.1.1 Let Ω be a domain in $\mathbb{C}$ and let $f : \mathbb{C} \to \widehat{\mathbb{C}}$ be any function. A point $\omega \in \mathbb{C}$ is called a period of f if

$$f(z + \omega) = f(z), \tag{10.2}$$

whenever both $z, z + \omega \in \Omega$.

Remark 10.1.1

1. By choosing Ω sufficiently small we can artificially create periods for f by merely choosing ω so that $\Omega \cap (\Omega + \omega) = \emptyset$. We are not interested in such cases and so require the domain Ω to satisfy the condition

$$\Omega + \omega = \Omega$$

whenever, ω is to be considered as a period for some function f on Ω. In other words, Ω is invariant under translation by ω (i.e., ω is period of the domain Ω).

2. In what follows we shall consider only a **meromorphic function**, though a number of observations we make below are valid for continuous functions as well.

3. Observe that 0 is a period of every function. If ω_1, ω_2 are periods of f then so is $\omega_1 + \omega_2$. Thus the set π_f of all periods of f forms an additive subgroup of $\mathbb{C}$. We shall call this the *period group* of f and denote it by π_f. If $\pi_f \neq (0)$ then we say f is a *periodic function.*

4. If z is a pole (or a zero) of f of order k then $z \pm \omega$ is also a pole (resp. a zero) of order k for all $\omega \in \pi_f$.

5. Let P_f denote the set of all poles of f. It may happen that $P_f \cap \pi_f \neq \emptyset$. Then it follows that $0 \in P_f$ and hence $\pi_f \subset P_f$. Otherwise, $\pi_f \subset \{z \in \mathbb{C} \ : \ f(z) = f(0)\}$. Hence, π_f is a closed discrete subset of Ω

6. Clearly, the properties of a periodic function are closely associated with the properties of its period group. However, we are interested presently in the properties of f vis-a-vis just one single period ω and not necessarily the whole group π_f. For this study, it does not matter even if ω itself is the integral multiple of another period of f.

7. A simple example of a function with period $\omega \neq 0$ is $e^{2\pi \imath z/\omega}$.

8. Let us now normalize and assume $\omega = 1$. The natural question is what are all periodic holomorphic (meromorphic) functions with period 1?

9. We would like to construct a periodic function with poles at all the integer points by considering sums such as

$$\sum_{-\infty}^{\infty} \frac{1}{z-n}. \tag{10.3}$$

However, there is the problem of convergence of such a sum. This can be settled by considering the so called *Eisenstein's sum* (see 8.3 in section 8.3). Alternatively, let us modify this and begin with the sum

$$f(z) = \sum_{-\infty}^{\infty} \frac{1}{(z-n)^2} \tag{10.4}$$

which is clearly convergent. It is easily checked that f is a singly periodic function with period 1. Integrating term-by-term yields,

$$g(z) = \frac{1}{z} + \sum_{n\neq 0}\left(\frac{1}{z-n} - \frac{1}{n}\right).$$

Again, g is a singly periodic function with simple poles at the integers. Indeed, one can check that $f(z) = \dfrac{\pi^2}{\sin^2 \pi z}$ (see exercise 8.3.1). From this it follows that f has no zeros and hence $\dfrac{1}{f(z)} = \dfrac{\sin^2 \pi z}{\pi^2}$ is a singly periodic holomorphic function. (This fact is not at all obvious from the series representation (10.4) of the function.)

10. It turns out that we need not go on looking for such constructions at all. Let us consider the function $\psi(z) = e^{2\pi\imath z/\omega}$ and $\Omega' = \psi(\Omega)$. Since ψ is an open mapping, it follows that Ω' is a domain. Now for any holomorphic (meromorphic) $g : \Omega' \to \mathbb{C}$ define $f(z) = g(e^{2\pi\imath z/\omega})$. Then f is a periodic holomorphic (meromorphic) function on Ω with ω as a period.

$$\begin{tikzcd} & \Omega' \arrow[d, "g"] \\ \Omega \arrow[ur, "\psi"] \arrow[r, "f"'] & \mathbb{C} \end{tikzcd}$$

Conversely, given any periodic function f with period ω we can define $g : \Omega' \to \mathbb{C}$ so that $f(z) = g \circ \psi(z)$ as follows. Given $u \in \Omega'$ pick up any z such that $\psi(z) = u$ and take $g(u) = f(z)$. The choice of z is unique up to an addition of $2m\pi\imath/\omega, m \in \mathbb{Z}$. But the value of $f(z)$ depends on u. Having picked up a z_0 over u_0, you can find a neighborhood U of z_0 such that $\psi|U$ is bijective onto a neighborhood of u_0 with an inverse ψ^{-1} which is holomorphic. It then follows that $g|U = f \circ \psi^1$ is also holomorphic. (Indeed $\psi^{-1}(u) = \frac{\omega}{2\pi\imath} \ln u$ for a branch of $\ln$.) Therefore, $g : \Omega' \to \mathbb{C}$ is holomorphic.

11. Suppose now that Ω is a parallel horizontal strip $c_1 < y < c_2$ and $\omega = 1$. Then Ω' is an annulus. Let g be holomorphic on Ω'. Let

$$g(w) = \sum_{n=-\infty}^{\infty} a_n w^n$$

be the Laurent expansion of g. Then

$$f(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi\imath z/\omega}$$

is the Fourier expansion of f in Ω which is a parallel strip. The coefficients a_n are given by

$$a_n = \frac{1}{2\pi\imath}\int_{|w|=r}\frac{g(w)}{w^{n+1}}dw = \frac{1}{\omega}\int_a^{a+\omega} f(z)e^{2n\pi\imath z/\omega}dz.$$

In the latter expression, the integrals are taken along any curve lying in Ω and joining a to $a+\omega$. This more or less brings the study of singly periodic functions to a close.

Exercise 10.1

1. We can take any continuous map $f : \mathbb{R} \to \mathbb{R}$ and define f to be periodic if $f(x+\omega) = f(x)$ for all x, where $\omega \neq 0$.

   (a) Show that the set π_f of all periods of f including 0 forms a closed additive subgroup of $\mathbb{R}$.

   (b) Show that π_f is discrete iff there is a least positive $\omega \in \pi_f$. In this case show that $\pi_f = \omega\mathbb{Z}$.

   (c) Show that π_f is either discrete or the whole of $\mathbb{R}$.

   (d) What can you say about f if $\pi_f = \mathbb{R}$?

   (e) Given any discrete subgroup G of $\mathbb{R}$ take any continuous function $g : \mathbb{R} \to \mathbb{R}$ which vanishes outside a closed interval and define $f(x) = \sum_{w\in G} g(x+w)$. Show that f is a continuous function and $G \subset \pi_f$. Can you get a function f so that $\pi_f = G$?

2. Extend the results in the above exercise 10.1.1 to $\mathbb{R}^2$ as follows: The definition of a periodic function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is the same. Ex.10.1.1(a) also extends ditto. However, for the rest of the exercise, you have to work harder.

   (a) Suppose $\omega_1, \omega_2 \in \pi_f$ are independent over $\mathbb{R}$. Let $\delta = max\{|\omega_1|, |\omega_2|\}$. Show that $\omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ intersects every ball in $\mathbb{R}^2$ of radius bigger than or equal to δ.

   (b) Suppose there exists $\omega \in \mathbb{R}^2$ such that $\mathbb{R}\omega \subset \pi_f$. Then show that the study of f reduces to the study of 1-variable function.

   So, in what follows assume that no line in $\mathbb{R}^2$ is completely contained π_f.

   (c) Show that for every line L passing through 0, $L \cap \pi_f$ is a discrete set.

(d) Show that there exist $\omega_1 \in \pi_f$ of least positive length. If $\omega_1\mathbb{Z} \neq \pi_f$, then show that there exists $\omega_2 \in \pi_f \setminus \omega_1\mathbb{Z}$ of least positive length. Also, show that $\omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z} = \pi_f$.

(e) Conclude that any closed subgroup of $\mathbb{R}^2$ is equal to one of the following: (i) (0), (ii) $\mathbb{Z}\omega$, (iii) $\mathbb{R}\omega$, (iv) $\mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$, (v) $\mathbb{R}\omega_1 \oplus \mathbb{Z}\omega_2$, (vi) $\mathbb{R}^2$.

3. Describe all domains Ω in $\mathbb{C}$ which are singly periodic, i.e., there exist $\omega \neq 0$ in $\mathbb{C}$ such that $\omega + \Omega = \Omega$.

10.2 Doubly Periodic Function

Once again, let Ω be a domain in $\mathbb{C}$ and f be a meromorphic function on Ω with period group $\pi_f \neq (0)$. Clearly $\pi_f = \mathbb{C}$ iff f is a constant. So, we shall now onward assume that f is a non constant periodic function (which is the same as saying π_f is a proper subgroup of $\mathbb{C}$.)

If you have gone through the exercises in the previous section then you know the various possibilities for π_f. Not all of those cases occur here. One can check which of the possibilities for π_f listed in Exercise 10.1.2e can actually occur for a meromorphic function. Instead, since it is much easier to work out the same directly for a meromorphic function, let us do this afresh.

Since $\pi_f \subset \{z \in \Omega \;:\; f(z) = f(0)\} \cup P_f$ and since f is meromorphic, it follows that π_f is discrete. Therefore there is a non zero complex number $\omega_1 \in \pi_f$ with $|\omega_1|$ minimum. It may now happen that $\pi_f \subset \mathbb{R}\omega_1$ in which case, you can show that $\pi_f = \omega_1\mathbb{Z}$. [For: given $\omega \in \pi_f$ you know $\omega = r\omega_1$ for some $r \in \mathbb{R}$. Then you can write $\omega = (m+s)\omega_1$, where $m \in \mathbb{Z}$ and $|s| \leq 1/2$. But then $s\omega_1 \in \pi_f$ with $|s\omega_1| < |\omega_1|$. Hence $s = 0$.] If this is the case, then f is called singly periodic and we have already discussed this in the previous section.

So, we now consider the case when $\pi_f \not\subset \mathbb{R}\omega_1$. Choose $\omega_2 \in \pi_f \setminus \mathbb{R}\omega_1$ such that $|\omega_2|$ is minimal. We then claim that $\pi_f = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$. [For: given any $\omega \in \pi_f$ since ω_1, ω_2 are linearly independent over $\mathbb{R}$, we can first of all write $\omega = r\omega_1 + s\omega_2, \;\; r, s \in \mathbb{R}$. We then choose $m, n \in \mathbb{Z}$ such that

$$\omega = (m\omega_1 + n\omega_2) + (r'\omega_1 + s'\omega_2); \quad |r'| \leq 1/2, |s'| \leq 1/2.$$

Since $|\omega_1| \leq |\omega_2|$ and that ω_1, ω_2 are linearly independent, using cosine rule, it follows that $|r'\omega_1 + s'\omega_2| < |\omega_2|$. This means $r' = 0 = s'$.] It follows that, in this case, π_f is actually the direct sum of the two infinite cyclic groups $\omega_j\mathbb{Z}, j = 1, 2$. Thus π_f is a free abelian group of rank 2. For future use we shall make a few definitions:

Definition 10.2.1 By a lattice in $\mathbb{R}^n$ we mean an additive subgroup of $\mathbb{R}^n$ of rank n.

Definition 10.2.2 We say $\{\omega_1, \omega_2\}$ is *good basis* for a lattice Γ in $\mathbb{R}^2$, if it is a basis and

$$|\omega_1| = min\{|\gamma| \ : \ 0 \neq \gamma \in \Gamma\} \text{ and } |\omega_2| = min\{|\gamma| \ : \ \gamma \in \Gamma \setminus \omega_1\mathbb{Z}\}.$$

As observed above such a basis always exists.

Remark 10.2.1

1. If $\{\omega_1, \omega_2\}$ is a good basis for Γ, so is $\{\pm\omega_1, \pm\omega_2\}$.
2. If $\{\omega_1, \omega_2\}$ is a good basis for Γ, then for any $t \in \mathbb{C}^*$, $\{t\omega_1, t\omega_2\}$ is a good basis for $t\Gamma$.
3. If $\{\omega_1, \omega_2\}$ is good basis for Γ and $\tau = \omega_2/\omega_1$, then $|\Re(\tau)| \leq 1/2$.

We leave each of the above remarks as exercises to the reader.

Functions with their period group equal to a lattice are called *elliptic functions.* You have to wait a little bit to know the justification for this name. At present this helps us to distinguish them from singly periodic functions.

From now onwards, we shall assume that $\Omega = \mathbb{C}$ and $\Gamma = \pi_f = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ where $\{\omega_1, \omega_2\}$ is a good basis for Γ.

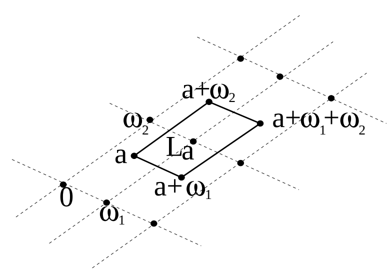


Fig. 48

We can choose any point $a \in \mathbb{C}$ and take the parallelogram L_a with vertices $a, a + \omega_1, a + \omega_2$ and $a + \omega_1 + \omega_2$ (see Fig. 48). It follows that the value of f is completely determined by the values of f on any such parallelogram.

Definition 10.2.3 We say $a \in \mathbb{C}$ is a good choice, if no zero or pole of f belongs to the boundary of L_a.

Remark 10.2.2 There are only a countable set of points in $\mathbb{C}$ which are **not** good choices.

Suppose now that $f : \mathbb{C} \longrightarrow \mathbb{C}$ is holomorphic. Then f is bounded on L_a and hence by periodicity, is bounded on the whole of $\mathbb{C}$. By Liouville's theorem f is a constant. Thus

Theorem 10.2.1 *A entire elliptic function is a constant.*

So, from now onwards, we focus on the larger class of functions, viz., meromorphic functions which are doubly periodic.

Lemma 10.2.1 Let P be the set of all poles of f inside L_a where a is a good choice. Then $\sum_{z\in P} Res_z(f) = 0$.

Proof: By residue theorem 5.7.1, the sum that we have to calculate is equal to $\dfrac{1}{2\pi\imath}\displaystyle\int_{L_a} f(z)dz$. By periodicity, the integrals on the opposite side of the parallelogram cancel out. ♠

Remark 10.2.3 Observe that f is elliptic implies so are f' and f'/f with $P_f = P_{f'} = P_{f'/f}$. Therefore the conclusion of the above lemma is applicable to f'/f.

Lemma 10.2.2 If $a \in \mathbb{C}$ is a good choice then the number of zeros of f inside L_a is equal to the number of poles of f inside L_a.

Proof: Remember that while counting zeros and poles we have to take them with their multiplicities. Then the difference of the above two quantities is equal to the sum total of the residues of f'/f. Clearly, the poles of f'/f are all simple and occur only at the zeros and poles of f. Therefore, we can apply the previous lemma to f'/f and conclude that the sum of the residues is zero. ♠

Remark 10.2.4

1. Let z be a zero of f of order k. By periodicity, the same holds for the points $z+\omega$, where ω varies over π_f. Thus, if Z is a complete set of representatives of the zero set of f modulo π_f, we define ord (f), *the order of f,* to be the sum of the orders of f at $z \in Z$. Clearly, given any good choice of a, Z can be chosen to be contained in L_a. In particular, it follows that Z is finite and hence $ord(f)$ is finite and is equal to the sum of the order of the poles of f over a complete set of representatives.

2. Further, it also follows that inside L_a, f assumes every value in $\hat{\mathbb{C}}$ exactly as many times as $ord(f)$. [For: we can apply the above discussion to the elliptic function $z \mapsto f(z) - w$ for any given $w \in \mathbb{C}$.]

Theorem 10.2.2 *Let $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ be the complete set of representatives of zeros and poles respectively. Then $\sum_j z_j - \sum_j w_j \in \pi_f$.*

Proof: Clearly it is enough to prove this for a complete set of representatives of zeros and poles contained in the interior of L_a for a good choice of a. Since

$$\sum_j z_j - \sum_j w_j = \int_{\partial L_a} \frac{zf'(z)}{f(z)} dz,$$

we shall show that this integral belongs to π_f.

Recall that for any two $z, w \in \mathbb{C}$, the line segment $[z, w]$ is parameterised by

$$t \mapsto (1-t)z + w.$$

Note that, since $f(a) = f(a + \omega_1)$, the curve $f \circ [a, a + \omega_1]$ is a closed curve not passing through 0. Hence the winding number $n_1 = \eta(f \circ [a, a+\omega_1]; 0)$ around 0 is a well-defined integer. Therefore (putting $z = w + \omega_2)$), we get

$$\begin{aligned} \int_{[a+\omega_2, a+\omega_1+\omega_2]} \frac{zf'(z)}{f(z)} dz &= \int_{[a,a+\omega_1]} \frac{(w+\omega_2)f'(w+\omega_2)}{f(w+\omega_2)} d(w+\omega_2) \\ &= \int_{[a,a+\omega_1]} \frac{wf'(w)}{f(w)} dw + n_1\omega_2. \end{aligned}$$

Therefore

$$\int_{[a+\omega_2, a+\omega_1+\omega_2]} \frac{zf'(z)}{f(z)} dz - \int_{[a,a+\omega_1]} \frac{zf'(z)}{f(z)} dz = n_1\omega_2 \in \pi_f.$$

Likewise we can show that the difference of the integrals on the other pair of opposite sides of L_a belongs to π_f. Since the sum of these gives the required integral on the boundary of L_a, we are done. ♠

We have not yet seen any doubly periodic function other than constants!

10.3 Weierstrass's Construction

We shall now take up the task of constructing doubly periodic functions. We fix a lattice $\Gamma = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ in $\mathbb{R}^2$ and for simplicity, consider periodic functions whose pole sets π_f coincide with Γ. The very first thing to observe is:

Theorem 10.3.1 *There are no doubly periodic functions with only simple poles at* Γ.

Proof: To see this, all we have to do is choose $a = (\omega_1 + \omega_2)/2$. (See Fig. 48.) Then L_a has no poles on its boundary. Indeed L_a will have exactly one point of Γ viz., $\omega_1 + \omega_2$ in its interior which happens to be a simple pole. Therefore the sum of residues cannot be zero, contradicting lemma 10.2.1. ♠

So, the simplest case we should seek is when the function has double poles at each point of Γ. This is precisely what Weierstrass had done.

Once again, we must be warned of the fact that double summations such as $\sum_{\gamma\in\Gamma} \frac{1}{(z-\gamma)^2}$ are not convergent. Hence, we are forced go in a round about manner. We begin with

$$f(z) = \sum_{\gamma\in\Gamma} \frac{1}{(z-\gamma)^3}. \tag{10.5}$$

Let us see why this is convergent. Observe that if $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is the linear transformation taking $\omega_1 \mapsto (1,0),\ \omega_2 \mapsto (0,1)$ then $T(m\omega_1 + n\omega_2) = (m,n)$. Therefore, $m^2 + n^2 \leq \|T\|^2\|m\omega_1 + n\omega_2\|^2$. Take $c = 1/\|T\|$. Given $R > 0$ choose $(m,n) \in \mathbb{R}^2 \setminus \mathbb{D}_R$. We then have,

$$|z - (m\omega_1 + n\omega_2)| \geq |m\omega_1 + n\omega_2| - |z| \geq c(m^2+n^2)^{1/2} - R$$

for all $|z| < R$.

Therefore,

$$\left|\frac{1}{(z - m\omega_1 - n\omega_2)^3}\right| \leq \frac{1}{(c(m^2+n^2)^{1/2} - R)^3} \leq \frac{M}{(m^2+n^2)^{3/2}}$$

Now the convergence of f follows from the lemma:

Lemma 10.3.1 $\displaystyle\sum_{m,n\in\mathbb{Z}} \frac{1}{(m^2+n^2)^\alpha}$ is convergent iff $\alpha > 1$.

Proof: By comparing with the area integral, the given sum is convergent iff $\displaystyle\iint_A \frac{r\,dr\,d\theta}{r^{2\alpha}}$ is convergent, where A is the annulus $|z| \geq 1$. But clearly, this area integral converges iff $\displaystyle\int_1^\infty \frac{dr}{r^{2\alpha-1}}$ converges iff $\alpha > 1$. ♠

Thus, $f(z)$ as in (10.5) is uniformly convergent on compact sets to a meromorphic function, which is clearly, Γ-periodic. This has poles of order three. To obtain a periodic function with poles of order 2, we now perform term-by-term integration of the sum minus the term $\frac{1}{z^3}$. After dividing out by $-1/2$, and adding $1/z^2$ back this yields

$$\boxed{\wp(z) = \frac{1}{z^2} + \sum_{\gamma\in\Gamma\setminus\{0\}} \left(\frac{1}{(z-\gamma)^2} - \frac{1}{\gamma^2}\right).} \tag{10.6}$$

This is called *Weierstrass's* $\wp$*-function* associated to the lattice Γ. (Read the symbol $\wp$ just like *pee*.) Observe that unlike under differentiation, periodicity is not preserved under integration, in general. So, we should justify the periodicity of (10.6).

Lemma 10.3.2 *Let f be an even function such that f' is periodic. Then f is periodic.*

Proof: For a fixed $\gamma \in \Gamma$ let $h(z) = f(z+\gamma) - f(z)$. Then clearly $h'(z) = f'(z+\gamma) - f'(z) = 0$. Therefore, $h(z) = c$. Putting $z = -\gamma/2$, we see that $c = 0$. Hence f is periodic. ♠

Remark 10.3.1 This then proves that $\wp$ is periodic and hence completes the construction that we started. Observe that the pole set of $\wp$ is contained the lattice Γ. So, the set of all Γ-periodic functions has some non constant functions in it. Our next task is to study this set.

10.4 Structure Theorem

As before, let $\Gamma = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$, be a lattice in $\mathbb{C}$. Let $\mathcal{M}_\Gamma$ be the set of all Γ-periodic meromorphic functions. Let us first of all sum up a few elementary facts about $\mathcal{M}_\Gamma$.

1. $\mathcal{M}_\Gamma$ is a field. In particular, if $R(t) = \dfrac{P(t)}{Q(t)}$ is a rational function of the variable t then $f \in \mathcal{M}_\Gamma \Longrightarrow R(f) \in \mathcal{M}_\Gamma$.
2. $f \in \mathcal{M}_\Gamma \Longrightarrow f' \in \mathcal{M}_\Gamma$.
3. Any non constant function $f \in \mathcal{M}_\Gamma$ has at most finitely many zeros and poles inside the fundamental domain.
4. The sum of the residues of f at all the poles inside L_a is zero.
5. Let $f \in \mathcal{M}_\Gamma$ be of order r. Then given any $c \in \mathbb{C}$, there are precisely r points $z \in \mathbb{C} \pmod{\Gamma}$ (counted with multiplicity) such that $f(z) = c$.

Let now $\wp$ be the Weierstrass' pee-function associated to the lattice Γ.

It follows that $\wp \in \mathcal{M}_\Gamma$ is one of the simplest doubly periodic functions with its pole set as Γ. It turns out that all other elements of $\mathcal{M}_\Gamma$ with this property can be got out of $\wp$ itself:

Theorem 10.4.1 *Let ϕ be any doubly periodic function with its pole set $P_f \subset \Gamma$. Then there exist two polynomial $h_1(t), h_2(t)$ such that $\phi(z) = h_1(\wp(z)) + \wp'(z)h_2(\wp(z))$. Moreover such an expression is unique.*

Proof: Write $\phi = \phi_1 + \phi_2$ as a sum of an even and an odd function. Then in the Laurent expansion of ϕ_1 around 0, the singular part will have terms involving only negative even powers of z :

$$\phi_1(z) = \frac{b_{2k}}{z^{2k}} + \cdots + \frac{b_2}{z^2} + f(z),$$

say, where f is an even holomorphic function. Clearly, then $\phi_1(z) - b_{2k}\wp^k(z)$ is an even periodic function with a pole of order $< 2k$. Hence, by induction, it follows that there is a polynomial h_1 of degree k such that $\phi_1(z) - h_1(\wp(z))$ is a holomorphic periodic function and hence is a constant. By adjusting this constant inside h_1, we get $\phi_1(z) = h_1(\wp(z))$. Now for the odd part, we observe that $\phi_2/\wp'$ is an even function and hence $= h_2(\wp(z))$ for some polynomial h_2. Putting these two together, we obtain

$$\phi(z) = h_1(\wp(z)) + \wp'(z)h_2(\wp(z)).$$

To see the uniqueness, suppose g_1, g_2 are two other polynomials such that $\phi(z) = g_1(\wp(z)) + \wp'(z)g_2(\wp(z))$. Then

$$h_1(\wp(z)) - g_1(\wp(z)) = \wp'(z)(g_2(\wp(z)) - h_2(\wp(z)).$$

LHS is an even function whereas RHS is an odd function. Therefore, both are identically zero. Since $(h_1 - g_1)(\wp(z)) = 0$ for infinitely many values of $\wp(z)$ this means $g_1 = h_1$ Similarly, $g_2 = h_2$. ♠

Finally we have the structure theorem:

Theorem 10.4.2 *Every element $f \in \mathcal{M}_\Gamma$ can be written in a unique way as*

$$f = R_1(\wp) + \wp' R_2(\wp) \tag{10.7}$$

where R_1, R_2 are rational functions.

Proof: This follows immediately from theorem 10.4.1 once we find a polynomial Q such that $Q(\wp)f$ has poles only in Γ. If, for a good choice of a, $a_1, \ldots, a_k$ are the poles of f inside some L_a enumerated with repetition to take care of the orders also, then take $Q(t) = \prod_{j=1}^{k}(t - \wp(a_j))$. Note that $Q(\wp) \in \mathcal{M}_\Gamma$ and $Q(\wp)f$ has no poles outside Γ. ♠

Remark 10.4.1 Note that the degree of the rational functions R_1, R_2 in (10.7) is determined by the order of poles of f. In particular, all doubly periodic functions with pole set Γ and having only poles of order two are linear combinations $a\wp + b$ where a, b are constants. Also, it follows that $(\wp')^2$ is a polynomial in $\wp$ of degree 3. Let us now compute this polynomial.

10.5 The Fundamental Relation

Let us introduce the notation

$$\sum_{\gamma}{}' := \sum_{\gamma\in\Gamma}{}' := \sum_{\gamma\in\Gamma\setminus\{0\}} .$$

We start with

$$\wp(z) = \frac{1}{z^2} + \sum_{\gamma}{}' \left(\frac{1}{(z-\gamma)^2} - \frac{1}{\gamma^2} \right). \tag{10.8}$$

Now consider the expansion around 0. It is clear that the singular part is $\frac{1}{z^2}$ and the constant term is zero. Since $\wp$ is an even function, the coefficients of all odd powers of z are also zero. Hence

$$\wp(z) = \frac{1}{z^2}\left(1 + \alpha z^4 + \beta z^6 + \cdots\right).$$

Also,

$$\wp'(z) = \frac{-2}{z^3} + 2\alpha z + 4\beta z^3 + \cdots .$$

Now a simple computation shows that

$$(\wp'(z))^2 - 4(\wp(z))^3 = \frac{-20\alpha}{z^2} - 28\beta + \cdots .$$

In particular, it follows that $(\wp'(z))^2 - 4(\wp(z))^3$ has poles of order two only and hence from the remark 10.4.1, is equal to $a\wp(z) + b$. The values of a, b are obtained by comparing the coefficients of $1/z^2$ and the constant term. We find that $a = -20\alpha, b = -28\beta$. Therefore,

$$(\wp'(z))^2 - 4(\wp(z))^3 + 20\alpha\wp(z) + 28\beta = 0.$$

It remains to compute the values of the constants α and β. Consider the difference function

$$\psi(z) = \wp(z) - \frac{1}{z^2} = \sum_{\gamma}{}' \left(\frac{1}{(z-\gamma)^2} - \frac{1}{\gamma^2} \right) = \alpha z^2 + \beta z^4 + \cdots$$

which is holomorphic near 0. Differentiating this twice and four times and then evaluating at $z = 0$ yields:

$$\psi^{(2)}(0) = 2\alpha = 6{\sum_\gamma}' \frac{1}{\gamma^4}; \quad \psi^{(4)}(0) = 24\beta = 120{\sum_\gamma}' \frac{1}{\gamma^6}.$$

In general, one denotes the so called *Eisenstein series of index k* by the notation

$$G_k := G_k(\Gamma) := {\sum_\gamma}' \frac{1}{\gamma^{2k}}. \tag{10.9}$$

We also have the traditional notation $g_2 = 60G_2, g_3 = 140G_3$. Thus we have proved

$$\boxed{\textbf{The Fundamental Relation: } (\wp')^2 = 4\wp^3 - g_2\wp - g_3.} \tag{10.10}$$

In other words, as z varies over $\mathbb{C}$, $(\wp(z), \wp'(z)) \in \mathbb{C} \times \mathbb{C}$ varies over the cubic curve of the form

$$\boxed{Y^2 = 4X^3 - g_2X - g_3.} \tag{10.11}$$

(Here do not confuse X, Y for real and imaginary parts of a complex variable z : X and Y are themselves complex variables.) Observe that (10.10) is a 1^{st}-order differential equation for the function $w = \wp(z)$ with the explicit solution:

$$z - z_0 = \int_{\wp(z_0)}^{\wp(z)} \frac{dw}{\sqrt{4w^3 - g_2w - g_3}} + \text{a constant}, \tag{10.12}$$

which is an elliptic integral. This is one of the justifications for calling $\wp$ an elliptic function, viz., $\wp$ is the inverse of a function given by an elliptic integral. [Suppose ϕ is a local inverse for $\wp$ i.e., $\wp(\phi(X)) = X$. Use chain rule and differentiate with respect to X to see that $\phi'(X)^2 = 4X^3 - g_2X - g_3$.]

Suddenly, it becomes very important for us to know how the curve (10.11) looks like. For instance, we may ask whether it is smooth or not. Recall that a curve $f(X, Y) = 0$ is smooth iff grad f does not vanish on the curve. Taking $f(X, Y) = Y^2 - 4X^3 - g_2X - g_3$, this just means that the polynomial $4X^3 - g_2X - g_3$ should not have any multiple roots. However, determining this seems to be an uphill task, if not impossible. So instead, we look out for more information from (10.10).

This leads us to the following observations:

(a) Since the $ord(\wp') = 3$, it follows that the equation $\wp' = 0$ has three solutions. It turns out that $\wp'$ has three distinct zeros (mod Γ) viz., $\omega_1/2, \omega_2/2$, and $(\omega_1 + \omega_2)/2$.

For, if τ denotes any one of these values, using periodicity and then oddness, we have, $\wp'(\tau) = \wp'(-\tau) = -\wp'(\tau)$. Therefore

$$\wp'(\tau) = 0, \quad \tau = \frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1+\omega_2}{2}.$$

There cannot be any other solution and each of these roots is simple. Let us put

$$\wp(\omega_1/2) = e_1, \quad \wp(\omega_2/2) = e_2 \ \ \& \ \ \wp((\omega_1+\omega_2)/2) = e_3. \tag{10.13}$$

(b) Observe that, since the derivative of $\wp(z) - e_j$ vanishes at these points, each of them has to be counted with multiplicity 2 as a solution of $\wp(z) = e_j$. Since the $ord\,\wp = 2$, it follows that e_1, e_2, e_3 are distinct. For if, say $e_1 = e_2$, then the function $\wp - e_1$ will have four zeros (counted with multiplicity) inside L_a.

(c) We can now prove the formula:

$$\boxed{\boxed{\wp^2 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3).}} \tag{10.14}$$

For, functions on both sides have poles of order 6 at 0 and zeros of order 2 at $\omega_1/2, \omega_2/2$ and $(\omega_1+\omega_2)/2$. Thus their quotient is doubly periodic and holomorphic and hence must be a constant. This constant is easily seen to be equal to 1 by comparing the coefficient of $\frac{1}{z^6}$. Thus we have successfully avoided direct factorization of the RHS of (10.11). Note down the method employed here for future use in establishing various identities of this type.

Remark 10.5.1

1. The uniqueness in theorem 10.4.1 actually implies that every relation between $\wp$ and $\wp'$ is a multiple of $q(X,Y) = Y^2 - 4X^3 - g_2X - g_3$. This fact may be seen as follows: Given $f(X,Y) \in \mathbb{C}[X,Y]$, such that $f(\wp, \wp') = 0$, using q we can replace all powers $Y^k, k \geq 2$ in f by a polynomial in X. That is to say $f(X,Y) = g(X,Y)q(X,Y) + h_1(X)Y + h_2(X)$ for some polynomials g, h_1, h_2. Since $f(\wp, \wp') = 0 = q(\wp, \wp')$, it follows that $h_1(\wp)\wp' + h_2(\wp) = 0$. That is $h_1(\wp)\wp' = -h_2(\wp)$ with one side odd and the other side even. Therefore, $h_1 = 0 = h_2$. Therefore, $f(X,Y) = g(X,Y)q(X,Y)$. In the fancy language of algebra, this can be expressed as an exact sequence

$$(0) \to (\alpha) \to \mathbb{C}[X,Y] \xrightarrow{T} \mathcal{M}'_\Gamma \to (0)$$

where T is the substitution mapping $T : \mathbb{C}[X,Y] \to M_\Gamma$ given by

$$X \mapsto \wp;\ Y \mapsto \wp'.$$

T is an algebra homomorphism with its image equal to the subring $\mathcal{M}'_\Gamma$ of $\mathcal{M}_\Gamma$ consisting of all Γ-periodic functions with their pole set inside Γ. The kernel of T is the principal ideal (q).

2. Just like $\wp$ itself was constructed by term-by-term integration of another elliptic function we can also integrate $\wp$. Since the sum of the residues of $\wp$ is zero, $\wp$ has a primitive. It is traditional to choose the function ζ so that ζ is an odd function and $\zeta' = -\wp$. It turns out that

$$\zeta(z) = \frac{1}{z} + {\sum_\gamma}' \left(\frac{1}{z-\gamma} + \frac{1}{\gamma} + \frac{z}{\gamma^2} \right). \tag{10.15}$$

Observe each term in this summation is got by integrating the corresponding term in (10.8) (except for $\frac{1}{z}$) and adjusting the sign. Of course, ζ is no longer elliptic. However, for each $\omega \in \Gamma$ since the derivative of $\zeta(z+\omega) - \zeta(\omega)$ vanishes, we get two constants η_1, η_2 such that

$$\zeta(z+\omega_j) - \zeta(z) = \eta_j, \quad j = 1, 2.$$

Observe that all poles of ζ are simple and inside Γ. Since the residue sum is equal to 1, by a simple integration along a parallelogram L_a which encloses 0, we obtain the so called *Legendre relation:*

$$\eta_1\omega_2 - \eta_2\omega_1 = 2\pi\imath. \tag{10.16}$$

3. Further integration of ζ is going to produce multi-valued map. By composing with the exponential map this can be converted into a single valued function σ with the property that

$$\frac{\sigma'(z)}{\sigma(z)} = \zeta(z). \tag{10.17}$$

It follows that σ has zeros which coincide with the pole set of ζ. Therefore Weierstrass's canonical product gives

$$\sigma(z) = z {\prod_\omega}' \left[\left(1 - \frac{z}{\omega}\right) \exp\left(\frac{z}{\omega} + \frac{z^2}{2\omega^2} \right) \right]. \tag{10.18}$$

4. Again, since
$$\frac{\sigma'(z+\omega_1}{\sigma(z+\omega_1)} = \zeta(z+\omega_1) = \zeta(z)+\eta_1 = \frac{\sigma'(z)}{\sigma(z)}+\eta_1$$
a simple integration shows that
$$\sigma(z+\omega_1) = K\sigma(z)\exp(\eta_1 z).$$
Check that σ is an odd function. Therefore by putting $z=-\omega_1/2$, we get $K = -\exp(\eta_1\omega_1/2)$. Similarly we can compute $\sigma(z+\omega_2)$ and we have
$$\sigma(z+\omega_j) = -\sigma(z)\exp\left(\eta_j\left(z+\frac{\omega_j}{2}\right)\right). \tag{10.19}$$

5. The function σ can be used to factorize any elliptic function. Indeed, let $f \in \mathcal{M}_\Gamma$ and let $a_1,\ldots,a_n$, and $b_1,\ldots,b_n$ be the zeros and poles of f inside a parallelogram. Then since $\sum_j a_j = \sum_j b_j$ it follows from (10.19) that the product
$$\frac{\prod_j \sigma(z-a_j)}{\prod_j \sigma(z-b_j)}$$
is an elliptic function having 'same' zeros and poles as f. Therefore we have
$$f(z) = c\frac{\prod_j \sigma(z-a_j)}{\prod_j \sigma(z-b_j)}$$
for some constant c.

10.6 The Elliptic Curve

Given a lattice $\Gamma \subset \mathbb{C}$ let us define an equivalence relation in $\mathbb{C}$ as follows:
$$z_1 \sim z_2 \quad \text{iff } z_1 - z_2 \in \Gamma.$$
The additive group structure on $\mathbb{C}$ passes down to the *orbit* set $\mathbb{C}/\Gamma$ of all equivalence classes so that the quotient map $q : \mathbb{C} \to \mathbb{C}/\Gamma$ is a (surjective) homomorphism with kernel equal to Γ. The topology on $\mathbb{C}$ gives rise to a topology on $\mathbb{C}/\Gamma$ the so called quotient topology, by the rule: a subset U of $\mathbb{C}/\Gamma$ is open iff $q^{-1}(U)$ is open in $\mathbb{C}$. Observe that the quotient topology on $\mathbb{C}/\Gamma$ can also be characterized by the property: for any topological space X, a function $f : \mathbb{C}/\Gamma \to X$ is continuous iff the composite $f\circ q : \mathbb{C} \to X$ is continuous. It follows that the group operations on $\mathbb{C}/\Gamma$ are continuous. This makes $\mathbb{C}/\Gamma$ into a topological group. Since the quotient space is the continuous image of the

compact set L_a, it also follows that $\mathbb{C}/\Gamma$ is compact. It is not very difficult to check that $\mathbb{C}/\Gamma$ is actually a Hausdorff space. Following the characterization of quotient topology, we make the following tentative definition:

Definition 10.6.1 A function $f : \mathbb{C}/\Gamma \to X$ is said to be holomorphic in an open set $U \subset \mathbb{C}/\Gamma$ iff $f \circ q : q^{-1}(U) \to X$ is holomorphic.

Remark 10.6.1 The above definition makes the topological group $\mathbb{C}/\Gamma$ into a complex 1-dimensional manifold also. With this structure, $\mathbb{C}/\Gamma$ is called a complex torus.

Remark 10.6.2 Recall that the extended complex plane $\widehat{\mathbb{C}}$ was the first non trivial example of a Riemann Surface. (See remark 3.8.1.3.) Complex tori are the next class of *Riemann surfaces*. It is not hard to see that q restricted to the interior of L_a is a homeomorphism for any a. Also, if $\{\omega_1, \omega_2\}$ is a basis for Γ then put

$$a_1 = a, a_2 = a + (\omega_1 + \omega_2)/3, a_3 = a + (\omega_1 + \omega_2)/2$$

and check that

$$\mathbb{C}/\Gamma = \cup_{j=1}^{3} q(L_{a_j}).$$

Writing q_j for $q|L_{a_j}$ it follows easily that $q_i^{-1} \circ q_j$ is a translation by $a_i - a_j$ on $L_{a_j} \cap L_{a_i}$ and hence are biholomorphic. This is another way to see that $\mathbb{C}/\Gamma$ is a Riemann Surface.

Remark 10.6.3 (This remark requires a bit more familiarity with certain topological results and may be skipped if you do not have them at present.) Let us write $T := T_\Gamma := \mathbb{C}/\Gamma$ and let us write $[z]$ for the equivalence class represented by $z \in \mathbb{C}$. We have seen that the mapping $f : z \mapsto (\wp(z), \wp'(z))$ defined on $\mathbb{C} \setminus \Gamma$ takes values on the cubic curve

$$E_\Gamma : Y^2 = 4X^3 - g_2 X - g_3$$

in $\mathbb{C} \times \mathbb{C}$. Because of the periodicity this defines a mapping $h : T \setminus \{[0]\} \to E_\Gamma$ so that $h[z] = f(z)$.

Since $f(z) \to \infty$ as z tends to a lattice point, it follows that h is a proper mapping. Therefore, it follows that h is surjective. To verify injectivity of h, let $(\wp(z_1), \wp'(z_1)) = (\wp(z_2), \wp'(z_2))$ with $[z_1] \neq [z_2]$. Since $\wp$ is an even function (and restricted to any L_a is two-to-one), it follows that $[z_2] = [-z_1]$. But then since $\wp'$ is an odd function, it follows that $\wp'(z_2) = -\wp'(z_1) = -\wp'(z_2) = 0$. But then $[z_2]$ is equal to one of the three classes $[\omega_1/2], [\omega_2/2], [(\omega_1 + \omega_2)/2]$. Now $[z_2] = [-z_1]$ implies $[z_2] = [z_1]$ which is a contradiction.

At all points where $\wp'(z) \neq 0$, $\wp$ is a local homeomorphism. Further at the three points where $\wp'(z) = 0$, since each of them is a simple root, it follows that $\wp'(z)$ is a local homeomorphism. Thus, f is a local homeomorphism and in particular, an open mapping. Therefore, so is h. Therefore h is a homeomorphism. With the complex differentiable structure on E_Γ coming from $\mathbb{C} \times \mathbb{C}$, it follows that h is a biholomorphic mapping. Indeed, $\hat{E}_\Gamma = E_\Gamma \cup \{\infty\}$ denotes the one-point compactification of the curve E, then we can extend h to a homeomorphism $\hat{h} : T \to \hat{E}_\Gamma$ by sending $[0] \mapsto \infty$. Using homogeneous coordinates for the curve E_Γ, we can identify $\hat{E}_\Gamma$ with a projective curve. It turns out that this curve is smooth even at ∞ and $\hat{h}$ is biholomorphic. Alternatively, we can simply use $\hat{h}$ and the manifold structure on T to think of $\hat{E}_\Gamma$ as a Riemann surface.

Remark 10.6.4 The bijection $\hat{h}$ can be used to transfer the additive group law onto $\hat{E}_\Gamma$. The geometric interpretation of this group operation is a very interesting topic that we cannot discuss here. (See the TIFR notes [Gu], for example.)

10.7 The Canonical Basis

Giving an ordered basis $\{\omega_1, \omega_2\}$ for a lattice is the same as giving an invertible 2×2 real matrix, $A \in GL(2; \mathbb{R})$: simply think of each $\omega_j = (\Re(\omega_j), \Im(\omega_j))$ as a real row-vector and take the matrix

$$A := \begin{bmatrix} \Re(\omega_2) & \Im(\omega_2) \\ \Re(\omega_1) & \Im(\omega_1) \end{bmatrix}.$$

(Just wait a minute before you object for the order in which in the rows are taken!) We shall identify the matrix A with the basis $\{\omega_1, \omega_2\}$.

The expression (10.8) for $\wp$, to begin with, depends on the choice of the basis elements of $\Gamma = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$. Thus we may treat $\wp$ as being defined on the space $\mathbb{C} \times GL(2; \mathbb{R})$. Likewise we may treat the two 'constants g_2, g_3 occurring in (10.10) as functions on $GL(2; \mathbb{R})$. To emphasis this fact, we temporarily introduce a modification in the notation:

$$\wp(z, A) = \wp(z); \quad g_2(A) = g_2; \quad g_3(A) = g_3, z \in \mathbb{C}, \ A \in GL(2, \mathbb{R}). \tag{10.20}$$

Let us study the behavior of these functions in the new variables A.

Remark 10.7.1

1. First of all observe that, since the convergence in (10.8) is absolute, we can permute the terms as we feel. Therefore, if we change the basis without changing the lattice, the summation in (10.8), remains unchanged. If we write

$$\omega_2' = a\omega_2 + b\omega_1; \ \omega_1' = c\omega_2 + d\omega_1$$

it follows that $\{\omega_1', \omega_2'\}$ is a basis for Γ iff the $a, b, c, d \in \mathbb{Z}$ and $ad - bc = \pm 1$, i.e, resulting matrix M is integral unimodular:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{Z}).$$

Therefore we have

$$\wp(z, MA) = \wp(z, A); \quad g_2(MA) = g_2(A); \quad g_3(MA) = g_3(A). \tag{10.21}$$

Thus, if $\mathcal{L}$ denotes the orbit space $GL(2;\mathbb{R})/GL(2;\mathbb{Z})$ (which can be identified with the set of all lattices in $\mathbb{R}^2$), then $\wp$ can be treated as a well defined function on $\mathbb{C} \times \mathcal{L}$.

2. Recall that the multiplication μ_t by a non zero number t defines a similarity on $\mathbb{C}$. If Γ is a lattice, so is $t\Gamma$. Thus $\mu_t : \mathbb{C} \to \mathbb{C}$ induces a biholomorphic mapping of the two tori $\hat{\mu}_t : \mathbb{C}/\Gamma \to \mathbb{C}/t\Gamma$. Conversely, using covering space theory, it is not hard to see that if two complex tori $\mathbb{C}/\Gamma, \mathbb{C}/\Gamma'$ are biholomorphic then there is a complex number $t \neq 0$ such that $\Gamma' = t\Gamma$.

   Thus we are lead to consider two lattices Γ, Γ' in $\mathbb{C}$ as equivalent if there is a complex number such that $\Gamma' = t\Gamma$. Let us denote the set of equivalence classes of lattices in $\mathbb{C}$ by $\mathcal{E}$.

3. How does the function $\wp$ behave under this equivalence? More specifically, what happens when the matrix A is changed to say $tA = \begin{bmatrix} t\omega_2 \\ t\omega_1 \end{bmatrix}$ for some $t \in \mathbb{C}^*$? We have

$$\begin{aligned} \wp(z, tA) &= \frac{1}{z^2} + \sum_{\gamma \in t\Gamma}{}' \left(\frac{1}{(z-\gamma)^2} - \frac{1}{\gamma^2} \right) \\ &= \frac{1}{t^2} \left(\frac{1}{z^2/t^2} + \sum_{\gamma \in \Gamma}{}' \left(\frac{1}{(z/t-\gamma)^2} - \frac{1}{\gamma^2} \right) \right) \\ &= \frac{1}{t^2} \wp(z/t, \Gamma)). \end{aligned}$$

In other words, we have the following homogeneity property:

$$\boxed{\boxed{\wp(tz, tA) = t^{-2}\wp(z, A).}} \tag{10.22}$$

The same holds for g_2 and g_3 as well, viz.,

$$\boxed{\boxed{g_2(tA) = t^{-4}g_2(A);\;\; g_3(tA) = t^{-6}g_3(A).}} \tag{10.23}$$

Therefore, it follows that the points (X, Y) on the cubic curve E_Γ are in 1-1 correspondence with the points on the cubic curve $E_{t\Gamma}$ under the mapping

$$(X, Y) \mapsto \left(\frac{X}{t^2}, \frac{Y}{t^3}\right). \tag{10.24}$$

4. Thus, in order to get functions which behave well under this equivalence, we must cook up certain homogeneous functions of degree zero out of the functions arising from $\wp$. For example discriminant $g_2^3 - 27g_3^2$ of the cubic $4X^3 - g_2X = g_3$ is homogeneous function of degree -6 in the variables (ω_1, ω_2). Therefore the quantity

   $$j = \frac{g_2^3}{g_2^3 - 27g_3^2}$$

   is a function which depends just on the ratio $\tau = \omega_2/\omega_1$. This is called the **j-function** which you may consider as defined on the upper-half plane. It plays a central role in the theory of elliptic curves. For details see for example [Gu].

   Yet another possibility is offered by the functions e_j considered in the previous section. This is the topic for the next section.

   However, in order to facilitate such a consideration, it is better to cut down the number of variables involved in $\wp$, especially in the $GL(2, \mathbb{C})$ part.

5. As a first step, we observe that the equivalence $\Gamma \sim t\Gamma$ at once allows us to choose the first basis element to be always equal to 1 as follows: Given any lattice $\Gamma = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ we can consider the lattice $\Gamma' = \frac{1}{\omega_1}\Gamma = \mathbb{Z} \oplus \tau'\mathbb{Z}$ where $\tau' = \omega_2/\omega_1$, which is equivalent to Γ. By further changing the sign of τ' if necessary, we may assume that $\Im(\tau') > 0$. We conclude that

Theorem 10.7.1 *There is a surjective mapping $\boldsymbol{H} \to \mathcal{E}$ given by $\tau \mapsto [\mathbb{Z} \oplus \tau\mathbb{Z}]$.*

Thus we can ow view $\wp$ as defined by on $\mathbb{C} \times \boldsymbol{H}$ via

$$(z, \tau) \mapsto (z, A)$$

where A is the matrix corresponding to the basis $(1, \tau)$.

6. Having started this game, we would like to cut down the domain of $\wp$ further, as much as possible. Let us see how.

   We can begin with a good basis $\{\omega_1, \omega_2\}$ for $\Gamma = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$. By remark 10.2.1.2, it follows that $\{1, \tau = \omega_2/\omega_1\}$ is also a good basis for $\Gamma' = \frac{1}{\omega_1}\Gamma$. As seen before, by replacing ω_2 by $-\omega_2$ if necessary, we can assume that $\Im(\tau) > 0$. It then follows that $|\tau| \geq 1$ and $|\Re(\tau)| \leq 1/2$. [For: we have $1 \leq |\tau| \leq |\tau \pm 1|$. Therefore, if $\tau_2 = x + \imath y$ we have, $1 \leq x^2 + y^2 \leq (x \pm 1)^2 + y^2$ which implies $|x| \leq 1/2$.] Moreover, if $\Re(\tau) = -1/2$ then we can replace it by $\tau + 1$ and assume that $\Re(\tau) = 1/2$. Finally, if $|\tau| = 1$ and if $\Re(\tau) < 0$, we perform one more change of lattice Γ to $\frac{1}{\tau}\Gamma$. In this, the first generator is again 1 whereas the second can be chosen to be $-1/\tau$ which will have its real part non negative. To sum up our observation:

Theorem 10.7.2 *Every lattice class in $\mathbb{C}$ has a representative with a basis $\{1, \tau\}$ where τ satisfies the following conditions:*
(i) $1 \leq |\tau|$
(ii) $\Im(\tau) > 0$.
(iii) $\frac{1}{2} < \Re(\tau) \leq \frac{1}{2}$.
(iv) If $|\tau| = 1$ *then* $\Re(\tau) \geq 0$.
Moreover such a basis is unique. (See the figure 48.)

Proof: It remains to prove the uniqueness. This just means that if τ, τ' are two such numbers such that $[\mathbb{Z} \oplus \tau\mathbb{Z}] = [\mathbb{Z} \oplus \tau'\mathbb{Z}]$, then we must show that $\tau = \tau'$. Let $t \in \mathbb{C}^*$ be such that $t(\mathbb{Z} \oplus \tau\mathbb{Z}) = \mathbb{Z} \oplus \tau'\mathbb{Z}$. Since in both groups, the least length of a non zero element is 1, it follows $|t| = 1$. This then also implies that the lattice $\mathbb{Z} \oplus \tau\mathbb{Z}$ is simply rotated about 0 to obtain the lattice $\mathbb{Z} \oplus \tau'\mathbb{Z}$. Therefore the second generators also have same modulus, i.e., $|\tau| = |\tau'|$. Since $\{t, t\tau\}$ and $\{1, \tau'\}$ are two basis for $t\Gamma$, there exist $a, b, c, d \in \mathbb{Z}$ such that $ad - bc = \pm 1$ and

$$t\tau = a\tau' + b; \ t = c\tau' + d.$$

Therefore,

$$\tau = \frac{a\tau' + b}{c\tau' + d}.$$

Comparing the imaginary parts on either side, it follows, first of all that $ad - bc = 1$ and then $\Im(\tau) = \Im(\tau')$. Combining this with the facts $\Re(\tau) \geq 0$ and $|\tau| = |\tau'|$, it follows that $\tau = \tau'$ ♠

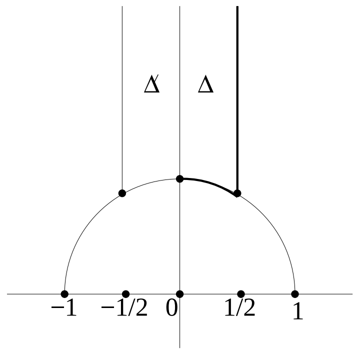


Fig. 48

Definition 10.7.1 We shall call such a basis given by the above theorem, the *canonical basis* for the lattice class $[\Gamma]$. The subset of $\mathbb{C}$ described by the above theorem will be called the fundamental region. The portion that lies on the right of $y-$axis is denoted by Δ and the portion on the left is denoted by Δ'. The fundamental region is actually equal to $\bar{\Delta} \cup \Delta'$. See the Fig. 48.

A Word of Caution Note that each of Δ, Δ' are open sets whereas, the fundamental region is not an open set. We caution you lest the usage 'fundamental region' confuse you.

Remark 10.7.2 Let us re-examine all the steps that we have taken in arriving at a canonical basis. Starting with a point $\tau \in \boldsymbol{H}$ we consider the class of the lattice $\mathbb{Z} \oplus \tau\mathbb{Z}$. The basis $\{1, \tau\}$ may not be a good basis. Suppose $\{\omega_1, \omega_2\}$ is a good basis and

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2, \mathbb{Z})$$

is such that $M(\tau, 1) = (\omega_2, \omega_1)$. This means that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \Re\tau & \Im\tau \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a\Re\tau + b & a\Im\tau \\ c\Re\tau + d & c\Im\tau \end{pmatrix} = \begin{pmatrix} \Re(\omega_2) & \Im(\omega_2) \\ \Re(\omega_1) & \Im(\omega_1) \end{pmatrix}.$$

This means

$$\tau' = \omega_2/\omega_1 = \frac{a\tau + b}{c\tau + d}. \tag{10.25}$$

Thus we see that the element τ' is got by effecting a Möbius transformation viz.,

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}, \quad a, b, c, d \in \mathbb{Z}.$$

Recall (see Exercises 6,7,8 in section 3.7) that the group of all Möbius transformations which map $\boldsymbol{H}$ onto itself is given by

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}$$

with $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. Amongst them those with $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$ form a subgroup which we shall call the modular group and denote it by $\mathcal{M}$.

Combining this observation with theorem 10.7.2 we get:

Theorem 10.7.3 *To every element $\tau \in \boldsymbol{H}$, there is a unique $\tau' \in \bar{\Delta} \cup \Delta'$ and a unique element $A \in \mathcal{M}$ such that $A\tau' = \tau$.*

Proof: Given $\tau \in \boldsymbol{H}$, we consider the lattice $\mathbb{Z} \oplus \tau\mathbb{Z} = \Gamma$. Let $\{\omega_1, \omega_2\}$ be a good basis for Γ such that $\Im(\omega_2/\omega_1) > 0$. Then as seen before it follows that the two basis are related via an element M of $SL(2, \mathbb{Z})$. Meanwhile as a Möbius transformation, we have also seen the effect of M on τ, i.e., $M\tau = \omega_2/\omega_1$. Now changing the lattice to $\frac{1}{\omega_1}\Gamma$ with the basis $\{1, \omega_2/\omega_1\}$ does not change this ratio. Already we have seen that $\tau' = \omega_2/\omega_1$ has the property that $|\tau'| \geq 1$, $\Im(\tau') > 0$ and $|\Re(\tau')| \leq 1/2$. If $\Re(\tau') = -1/2$ then we are replacing τ' by $\tau' + 1$ This change of basis is effected by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. As a Möbius transformation also τ' is mapped to $\tau' + 1$. Finally if $|\tau'| = 1$ and $\Re(\tau') < 0$, we replace the lattice by $\frac{1}{\tau'}\Gamma'$ and choose the basis $\{1, -1/\tau'\}$. This change can be effected by the Möbius transformation $z \mapsto -1/z$ corresponding to the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The rest of the proof follows from theorem 10.7.2 ♠

Remark 10.7.3 This is the reason why $\bar{\Delta} \cup \Delta'$ is called a fundamental region for the action of $\mathcal{M}$ on $\boldsymbol{H}$. (So, now you can compare your answer to Ex. 8 of section 3.7 with this.)

10.8 The Modular Function λ

Given any τ with $\Im(\tau) > 0$, we can consider the lattice class of $\mathbb{Z} \oplus \tau\mathbb{Z}$ with the basis $\{1, \tau\}$. We can then treat $\wp$ as a function of (z, τ).

Recall that the three roots e_1, e_2, e_3 of the cubic (10.11) were defined by

$$e_1 = \wp\left(\frac{1}{2}, \tau\right), e_2 = \wp\left(\frac{\tau}{2}, \tau\right), e_3 = \wp\left(\frac{1+\tau}{2}, \tau\right).$$

In this way, each of e_j can be considered as a holomorphic function of the variable τ on the upper-half plane $\boldsymbol{H}$. Since e_j are distinct, it follows that the function

$$\lambda(\tau) = \frac{e_3 - e_2}{e_1 - e_2} \tag{10.26}$$

is a holomorphic function which never takes the value 0 or 1. We would like to study the behavior of λ especially with respect to the action of the modular group on $\boldsymbol{H}$. From (10.8), we have,

$$\wp(z, \tau) = \frac{1}{z^2} + \sum_{(m,n)\neq(0,0)} \left(\frac{1}{(z-m-n\tau)^2} - \frac{1}{(m+n\tau)^2}\right). \tag{10.27}$$

Therefore,

$$e_1(\tau) = \wp(1/2, \tau) = 4 + \sum_{(m,n)\neq(0,0)} \left(\frac{1}{(\frac{1}{2}-m-n\tau)^2} - \frac{1}{(m+n\tau)^2}\right).$$

Since the sum is absolutely convergent and each term t_n in the summation has the property that $t_n(\bar{\tau}) = \overline{t_n(\tau)}$, it follows that $e_1(\bar{\tau}) = \overline{e_1(\tau)}$.

On the other hand, $e_1(-\tau) = e_1(\tau)$, since the terms in the summation simply get permuted. Therefore, if τ is purely imaginary, then it follows that

$$e_1(\tau) = e_1(-\tau) = e_1(\bar{\tau}) = \overline{e_1(\tau)}.$$

Therefore e_1 takes real values on the positive imaginary axis. The same holds for e_2 as well as e_3. In particular,

Lemma 10.8.1 λ takes real values on the imaginary axis.

Also note that

$$\begin{aligned}(e_3 - e_2)(\tau) &= \sum_{m,n} \left(\frac{1}{[(m+\frac{1}{2})+(n+\frac{1}{2})\tau]^2} - \frac{1}{[m+(n+\frac{1}{2})\tau]^2}\right); \\ (e_1 - e_2)(\tau) &= \sum_{m,n} \left(\frac{1}{[(m+\frac{1}{2})+n\tau]^2} - \frac{1}{[m+(n+\frac{1}{2})\tau]^2}\right).\end{aligned} \tag{10.28}$$

From this it is easy to check that $(e_3 - e_2)(\tau + 2) = (e_3 - e_2)(\tau)$ and $(e_1 - e_2)(\tau + 2) = (e_1 - e_2)(\tau)$. Therefore,

$$\boxed{\boxed{\lambda(\tau + 2) = \lambda(\tau).}} \tag{10.29}$$

Let us take a minute to see what (10.29) actually means. Let Λ denote the *congruence subgroup* of all Möbius transformation A of the form

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ mod } 2.$$

Consider the two basis $B_1 = \{1, \tau\}$ and $B_2 = \{1, \tau'\}$ where, $\tau' = A\tau = (a\tau + b)/(c\tau + d)$. From (10.22), we have,

$$\begin{aligned} (c\tau + d)^{-2}\wp(1/2, B_2) &= \wp((c\tau + d)/2, (c\tau + d)B_2) = \wp((c\tau + d)/2, B_1) \\ &= \wp((2m\tau + 2n + 1)/2, B_1) = \wp(1/2, B_1). \end{aligned}$$

Note that the justification for the last step is $m\tau + n$ is in the group spanned by B_1. Similarly, it can be checked that

$$(c\tau + d)^{-2}\wp(\tau'/2, B_2) = \wp(\tau/2, B_1); \quad (c\tau + d)^{-2}\wp((1 + \tau')/2, B_2) = \wp((1 + \tau)/2, B_1).$$

Therefore each of the e_j gets multiplied by $(c\tau + d)^{-2}$ and hence λ remains unchanged. Thus we have proved:

Lemma 10.8.2 $\lambda(A\tau) = \lambda(\tau)$ for all $A \in \Lambda, \tau \in \boldsymbol{H}$.

There is an exact sequence of group homomorphisms

$$(1) \longrightarrow \Lambda \xrightarrow{q} SL(2, \mathbb{Z}) \longrightarrow SL(2; \mathbb{Z}/2\mathbb{Z}) \longrightarrow (1)$$

where q is defined by the reduction mod 2 homomorphism $\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$. The following six elements T_j of $SL(2, \mathbb{Z})$

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; T_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; T_3 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}; \tag{10.30}$$

$$T_4 = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}; T_5 = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}; T_6 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \tag{10.31}$$

map onto the six elements of $SL(2;\mathbb{Z}/2\mathbb{Z})$ under q. All this just means that Mod 2, the above six elements form a complete list of mutually incongruent elements modulo the subgroup Λ, in the group $SL(2\mathbb{Z})$. Thus we have the left-coset decomposition of $SL(2,\mathbb{Z})$:

$$SL(2,\mathbb{Z}) = \coprod_{j=1}^{6} T_j\Lambda$$

Let us put

$$V = \coprod_{j=1}^{6} T_j\Delta; \quad V' = \coprod_{j=1}^{6} T_j\Delta'.$$

It follows from theorem 10.7.2 that

$$\boldsymbol{H} = \coprod_{B\in\Lambda} B(\bar{V}\cup V').$$

The following familiar picture shows $\bar{V}\cup\bar{V'}$.

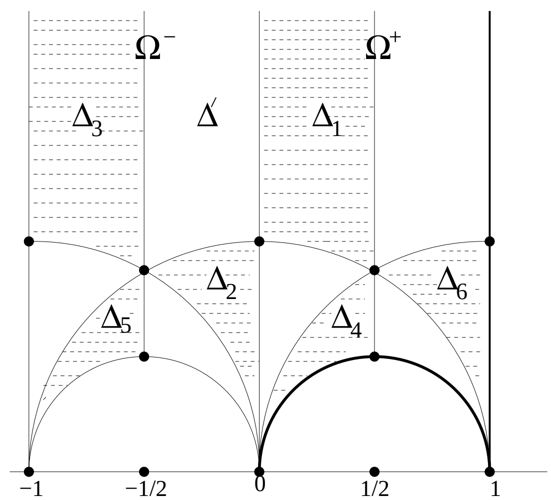


Fig. 49

The five circular arcs are parts of the five circles $|z\pm 1/2| = 1/2; |z| = 1; |z\pm 1| = 1$. Verify that $\Delta_j := T_j(\Delta), j = 1,2,\ldots,6$ are as indicated by shaded portions. The set V is contained in $\Omega^+\cup\Omega^-$ where

$$\Omega^+ = \{z\in\boldsymbol{H} \ : \ |z-1/2| > 1/2, \ \& \ \ 0 < \Re(z) < 1\};$$

$$\Omega^- = \{z\in\boldsymbol{H} \ : \ |z+1/2| > 1/2, \ \& \ -1 < \Re(z) < 0\}.$$

Observe that $V\subset\Omega^+\cup\Omega^-$ whereas some portions of V' are going out of this region. We can now trade portions of V' lying outside $\Omega^+\cup\Omega^-$ with those inside by simply replacing $T_j's$ by an appropriate representative modulo Λ viz., replace T_j respectively by

$$T_1' = T_1; T_2' = T_2; T_3' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; T_4' = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}; T_5' = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}; T_6' = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

It follows easily that

$$\bar{V} \cup \cup_j T_j'(\bar{\Delta}') = \bar{\Omega}^+ \cup \bar{\Omega}^-.$$

Therefore, it follows that $\Lambda(\bar{\Omega}^+ \cup \bar{\Omega}^-) = \boldsymbol{H}$. Finally the unwanted boundary parts can also be traded easily. Thus:

Theorem 10.8.1 *For every $\tau \in \boldsymbol{H}$, there exists a unique $A \in \Lambda$ and a unique $\tau' \in \bar{\Omega}^+ \cup \Omega^-$ such that $A\tau' = \tau$.*

Thus the study of the function λ is reduced to the region $\bar{\Omega}^+ \cup \bar{\Omega}^-$ and understanding the behavior of λ under the Möbius transformations corresponding to $T_j's$. The following table gives six representative matrices of $\mathcal{M}/\Lambda$, the corresponding Möbius transformation, their values on the elements $\imath, e^{\pi\imath/3}$ and ∞ and their effect on the Modular function λ.

T_1	T_2	T_3	T_4	T_5	T_6
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$
τ	$-\frac{1}{\tau}$	$\tau - 1$	$\frac{1}{1-\tau}$	$\frac{\tau-1}{\tau}$	$\frac{\tau}{1-\tau}$
$\imath$	$\imath$	$\imath - 1$	$\frac{1+\imath}{2}$	$1+\imath$	$\frac{\imath-1}{2}$
$e^{\pi\imath/3}$	$e^{2\pi\imath/3}$	$e^{2\pi\imath/3}$	$e^{\pi\imath/3}$	$e^{\pi\imath/3}$	$e^{2\pi\imath/3}$
∞	0	∞	0	1	-1
$\lambda(\tau)$	$1-\lambda(\tau)$	$\frac{\lambda(\tau)}{\lambda(\tau)-1}$	$\frac{1}{1-\lambda(\tau)}$	$\frac{\lambda(\tau)-1}{\lambda(\tau)}$	$\frac{1}{\lambda(\tau)}$

Using (10.22) and arguing as in the proof of lemma 10.8.2, it follows that

$$\lambda(T_2(\tau)) = \lambda\left(-\frac{1}{\tau}\right) = 1 - \lambda(\tau).$$

Since replacing τ by $\tau + 1$ does not change $\wp$ but interchanges e_3 and e_2, it follows that

$$\lambda(T_3\tau) = \lambda(\tau - 1) = \frac{\lambda(\tau)}{\lambda(\tau) - 1}.$$

Other relations can be deduced from these:

$$\lambda(T_4\tau) = \lambda\left(\frac{1}{1-\tau}\right) = 1 - \lambda(\tau - 1) = 1 - \frac{\lambda(\tau)}{\lambda(\tau) - 1} = \frac{1}{\lambda(\tau) - 1}.$$

$$\lambda(T_5\tau) = \lambda\left(\frac{\tau-1}{\tau}\right) = \lambda\left(-\frac{1}{\tau}+1\right) = \frac{\lambda(-1/\tau)}{\lambda(-1/\tau)-1} = \frac{1-\lambda(\tau)}{1-\lambda(\tau)-1} = \frac{\lambda(\tau)-1}{\lambda(\tau)}.$$

$$\lambda(T_6\tau) = \lambda\left(-\frac{1}{T_5\tau}\right) = 1 - \frac{\lambda(\tau)-1}{\lambda(\tau)} = \frac{1}{\lambda(\tau)}.$$

Further, since $\lambda(-\tau) = \overline{\lambda(\tau)}$, we can simply concentrate on $\bar{\Omega}$, and ignore Ω^-.

We have already proved that λ maps the positive imaginary axis to the real line. By the relation $\lambda(\tau+1) = \lambda(\tau-1) = \frac{\lambda(\tau)-1}{\lambda(\tau)}$, it follows that the line $\Re(\tau) = 1$ is also mapped to the real line. The circular boundary part $|z-1/2| = 1/2$ is mapped onto the line $\Re(\tau) = 1$ by T_4. Hence λ sends this boundary part also into the real line.

We claim:

Lemma 10.8.3

(i) $\lambda(\tau) \to 0$, as $\Im(\tau) \to \infty$ uniformly with respect to the real part of τ.

(ii) $\lambda(\tau) \to 1$, as $\tau \to 0$ inside Ω

(iii) $\lambda(\tau) \to \infty$, as $\tau \to 1$ inside Ω.

(iv) $\lambda(\tau)e^{-\pi\imath\tau} \to 16$, as $\Im(\tau) \to \infty$.

Proof: Recall that (see exercise 8.3.1 from section 8.3):

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{-\infty}^{\infty} \frac{1}{(z-m)^2}.$$

Now sum the two series (10.28) with respect to m to get

$$e_3 - e_2 = \pi^2 \sum_{-\infty}^{\infty} \left(\frac{1}{\cos^2 \pi(n-\frac{1}{2})\tau} - \frac{1}{\sin^2 \pi(n-\frac{1}{2})\tau} \right);$$

$$e_1 - e_2 = \pi^2 \sum_{-\infty}^{\infty} \left(\frac{1}{\cos^2 \pi n\tau} - \frac{1}{\sin^2 \pi(n-\frac{1}{2})\tau} \right).$$

Note that the convergence is uniform in $\Im(\tau) \geq \delta > 0$. Therefore, we can take the limit term-by-term and conclude that $(e_3 - e_2) \to 0$ and $(e_1 - e_2) \to \pi^2$ (contribution from the term $n = 0$) as $\Im(\tau) \to \infty$. This proves (i). To prove (ii) we use the fact that the Möbius transformation $\tau \mapsto -1/\tau$ maps the region $B_r(0) \cap \Omega$ inside $\Im(z) > \delta$ for some $\delta > 0$. Also, $\lambda(-\frac{1}{\tau}) = 1 - \lambda(\tau)$ and hence $\lambda(re^{\imath\theta}) = 1 - \lambda(e^{-\imath\theta/r}) \to 1$ as $r \to 0$. To prove (iii) use the fact that the Möbius transformation $\tau \mapsto \frac{\tau}{1-\tau}$ maps the region $B_r(1) \cap \Omega$ inside $\Im(z) > \delta$ for some $\delta > 0$ and the relation $\lambda(\frac{\tau}{\tau-1}) = \frac{1}{\lambda(\tau)}$. To prove (iv), we rewrite $e_3 - e_2$ in terms of $e^{\pi\imath\tau}$: Since

$$\frac{1}{\cos^2 z} - \frac{1}{\sin^2 z} = 4e^{2\imath z} \left(\frac{1}{(1+e^{2\imath z})^2} + \frac{1}{(1-e^{2\imath z})^2} \right)$$

it follows that $e_3 - e_2 = 4\pi^2 \sum_{n=-\infty}^{\infty} t_n$, where

$$t_n = e^{(2n-1)\imath\pi\tau}\left(\frac{1}{(1+e^{(2n-1)\imath\pi\tau})^2} + \frac{1}{(1-e^{(2n-1)\imath\pi\tau})^2}\right).$$

Check that for $n \neq 0, 1$,

$$\lim_{\Im(\tau)\to\infty} t_n e^{-\pi\imath\tau} = 0;$$

whereas for $n = 0$ and 1, these limits are both equal to 2. Since $e_1 - e_2 \to \pi^2$, anyway this proves (iv). ♠

Theorem 10.8.2 *The modular function λ defines a bijective mapping of Ω onto the upper-half plane.*

Proof: We plan to prove this using the argument principle. Fix $w \in \boldsymbol{H}$. Choose $R > 0$ such that $|w| < R$. By lemma 10.8.3, we can choose $r > 0$ such that for all $0 < \rho \leq r$ $\lambda(B_\rho(1) \cap \Omega^+) \cap \mathbb{D}_R = \emptyset$. Since w is not on the real line we can (and will) re-choose $r > 0$ such that w is outside the two discs $|z| \leq r$ and $|z - 1| \leq r$. We then choose $r > s > 0$ such that λ maps the region $\mathbb{D}_s \cap \Omega$ inside $B_r(1)$. Finally we choose $\delta >> 0$ such that the region $\Im(\tau) > \delta$ is mapped inside $\mathbb{D}_s$.

Let p_1, p_2 be the point of intersection of the circle $|z - 1/2| = 1/2$ with the circles $|z| = s$ and $|z - 1| = r$ respectively. Fix $y_0 > \delta$. Consider the closed curve γ in Ω consisting of

(i) $L_1 = [1 + \imath y_0, \imath y_0]$ followed by

(ii) $L_2 = [\imath y_0, \imath s]$, followed by

(iii) L_3, the arc of the circle $|z| = s$ from $\imath s$ to p_1 followed by

(iv) L_4, the arc of the circle $|z - 1/2| = 1/2$ from p_1 to p_2, followed by

(v) L_5, the arc of the circle $|z - 1| = r$ from p_2 to $1 + \imath r$, and finally

(vi) $L_6 = [1 + \imath r, 1 + \imath y_0]$. (See Fig.49)

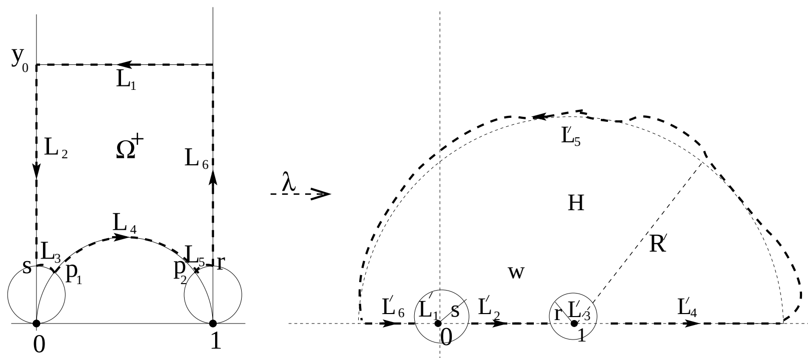


Fig. 49

Let the image of the arcs L_j under λ be denoted by L'_j. We then know that

(i) L'_1 is some curve completely inside the disc $B_s(0)$;

(ii) L'_2 is contained in the line segment $[0, 1]$;

(iii) L'_3 is contained in the disc $B_r(1)$;

(iv) L'_4 is contained in the ray $[1, \infty)$;

(v) L'_6 is contained in the ray $(-\infty, 0]$.

What happens to L'_5? Of course L'_5 is contained in the upper half plane and its two end points lie on the real line. By the choice of r, L'_5 lies in the complement of the disc $\mathbb{D}_R$. Check that L_5 is the image of the segment $L = [\imath/r, 1 + \imath/r]$ under the Möbius transformation $\tau \mapsto 1 - 1/\tau$. From (iv) of lemma 10.8.3, $\lambda(L)$ is a curve very close to the circle $|z| = 16e^{-\pi/r}$. Since $\lambda(1 - 1/\tau) = 1 - 1/\lambda(\tau)$, it follows that L'_5 is a curve very close to the image of the above circle under the transformation $\tau \mapsto 1 - 1/\tau$ which the circle $|z - 1| = e^{\pi/r}/16 =: R'$.

From these considerations, it follows that the winding number $\eta(\lambda(\gamma); w)$ is zero if w is in the lower half plane and 1 if it is in the upper-half plane. (This is where you may employ lemma 5.6.2 of section 5.5 for a rigorous argument.) From this, we conclude that λ does not assume any value in the lower half plane and assumes every value in the upper-half plane exactly once, i.e., $\lambda : \Omega^+ \to \boldsymbol{H}$ bijective. ♠

Remark 10.8.1

1. By reflection principle, λ maps Ω^- onto lower-half-plane. Thus we have $\lambda : \bar{\Omega}^+ \cup \Omega^- \to \mathbb{C} \setminus \{0, 1\}$ which is bijective on either side of the imaginary axis to either side of the real axis. It remains to see what happen on the three boundary pieces.

2. We claim that λ is monotone on each boundary piece. For if not, then there would be a point at which its derivative would vanish. (See Exercise 3.3.5.) But then,

we know that the mapping would be $n - to - 1$ in a neighborhood of that map for $n \geq 2$, which is absurd.

3. Since each of the boundary piece is mapped inside a different component of $\mathbb{R} \setminus \{0, 1\}$, we conclude that λ is one-to-one mapping on $\bar{\Omega}^+$. Since the interior of the region is strictly mapped inside the upper-half plane it follows that $\lambda : \bar{\Omega}^+ \to \bar{H} \setminus \{0, 1\}$ is a proper mapping. From this we see that it is also onto.

4. Let us denote by U the interior of $\bar{\Omega}^+ \cup \Omega^-$. It follows that $\lambda : U \to \mathbb{C} \setminus (-\infty, 0] \cup [1, \infty)$ is biholomorphic.

5. Check that
(i) $\lambda^{-1}(\mathbb{C} \setminus ((-\infty, 0] \cup [1, \infty))) = \coprod_{A \in \Lambda} AU$, (a disjoint union);
(ii) $\lambda^{-1}(\mathbb{C} \setminus (-\infty, 1]) = \coprod_{A \in \Lambda} A(T_4 U)$;
(iii) $\lambda^{-1}(\mathbb{C} \setminus [0, \infty)) = \coprod_{A \in \Lambda} A(T_5 U)$.
In each case, λ restricted to each component defines a homeomorphism (in fact a biholomorphic mapping) of the component onto the respective open set in $\mathbb{C} \setminus \{0, 1\}$. Thus, in the terminology of definition 10.9.2, we have just proved:

Theorem 10.8.3 *The modular function $\lambda : \boldsymbol{H} \to \mathbb{C} \setminus \{0, 1\}$ is a covering projection.*

10.9 Picard Theorems

In this section, our aim is to give a proof of Big Picard theorem using the modular function as per in the original proof due to Picard. This needs us to construct a holomorphic mapping $g : \Omega \to \boldsymbol{H}$ so that $\lambda \circ g = f$, where f is a given map on Ω with certain properties, which turns out to be purely a topological problem. Due to the importance of the concept involved, viz., the lifting properties of covering projection, we shall first of all briefly treat this topic. The actual proof of Picard theorem then comes out very quickly, with the help of Montel's theorem and Casorati-Weierstrass's theorem. The proof runs similar to the one given in [N] except for the fact that we do not need Schottky's theorem. However, we must be apologetic about this, since Picard theorems indeed belong to a topic where one discusses the values of holomorphic functions such as Bloch's and Schottky's results.

Definition 10.9.1 Let $p : E \to B$ be a surjective continuous map of topological spaces. We say an open subset V of B is *evenly covered* by p, if $p^{-1}(V)$ is a disjoint union of open subsets of E :

$$p^{-1}(V) = \coprod_i U_i$$

where, each U_i is mapped homeomorphically onto V by p. If there is a family of open sets $\{U_\alpha\}$ in B such that $B = \cup_\alpha U_\alpha$ and each U_α is evenly covered by p, then we say that, p is a *covering projection*; the space E is called a *covering space* of B.

Remark 10.9.1

(a) We also say, E is the *total space* and B is the *base space* of the covering projection p.

(b) Every covering projection is a local homeomorphism. (Recall that, p is a local homeomorphism if $\forall\ y \in E$, there exists an open neighborhood U of y such that $f|_U$ is a homeomorphism of U onto an open subset of B). Indeed, E and B share all local topological properties of each other. For example, B is locally compact (respectively, locally connected, locally path connected) iff so is E.

(c) Every local homeomorphism is an open map and so is every covering projection. In general, given a map $f : X \to Y$, and a point $y \in Y$, we call the set $f^{-1}(y)$, *fiber of f over y*.

(d) If f is a local homeomorphism then the fibers of f are discrete, i.e., the subspace topology on $f^{-1}(y)$ is discrete. In particular, the fibers of a covering projection are discrete. This fact is going to play a very important role in what follows.

Example 10.9.1

(a) Any homeomorphism is a covering projection.

(b) A typical example of a covering projection is already familiar to you, viz., $\exp : \mathbb{R} \to \mathbb{S}^1$. For any fixed $\theta : 0 \leq \theta < 2\pi$, if we consider $U = \mathbb{S}^1 \setminus \{e^{\imath\theta}\}$, then $(exp)^{-1}(U)$ is the union of disjoint intervals, $\theta + 2n\pi < t < \theta + (2n+2)\pi$. Restricted to any of these intervals, exp is a homeomorphism. More interesting fact is that the map $exp : \mathbb{C} \to \mathbb{C}^*$ is also a covering projection. Verify this.

(c) In a similar way, it is not hard to see that the map $z \mapsto z^n$ defines a covering projection of $\mathbb{C}^\star$ onto itself. Here, n is any positive integer, and $\mathbb{C}^\star$ denotes the space of non zero complex numbers. This map restricted to the subspace $\mathbb{S}^1$ of complex numbers of modulus 1 defines a covering projection of $\mathbb{S}^1$ onto itself.

Definition 10.9.2 Let $p : E \to B$ be a covering projection and $f : Y \to B$ be a continuous map. By a lift (or a section) of f we mean a continuous map $\tilde{f} : Y \to E$ such that $p \circ \tilde{f} = f$.

Remark 10.9.2 Observe that, in addition, if p and f are holomorphic mappings then any lift $\tilde{f}$ of f is holomorphic.

Theorem 10.9.1 *Let $p : E \to B$, be a covering projection, Y be any connected space and $f : Y \to B$ be any map. Let g_1 and g_2 be any two lifts of f, such that, for some point $y \in Y$, $g_1(y) = g_2(y)$. Then $g_1 = g_2$.*

Proof: Let $Z = \{y \in Y : g_1(y) = g_2(y)\}$. It is given that, Z is nonempty. Thus, if we show that, Z is open and closed then from the connectivity of Y, it follows that $Z = Y$, i.e., $g_1 = g_2$.

Let $y \in Z$ and let V be an evenly covered open neighborhood of $f(y)$ in B. Let U be an open subset of E mapped homeomorphically onto V by p and let $g_1(y) = g_2(y) \in U$. Choose W, an open neighborhood of y in Y such that, $g_j(W) \subset U$ for $j = 1, 2$. Then, $p \circ g_1(z) = f(z) = p \circ g_2(z)\ \forall\ z \in W$. Since $p|_U$ is injective, this implies that, $g_1(z) = g_2(z)\ \forall\ z \in W$ and hence $W \subset Z$. Hence, Z is open.

So, let z be a point in Y such that, $g_1(z) \neq g_2(z)$. Let V be an evenly covered open neighborhood of $p \circ g_1(z) = p \circ g_2(z)$. There exists an open neighborhood U_j of $g_j(z)$ on which p is a homeomorphism and such that, $U_1 \cap U_2 = \emptyset$. Let W be an open neighborhood of z such that, $g_j(W) \subseteq U_j,\ j = 1, 2$. Then it follows that, W is an open neighborhood of z not intersecting Z. Hence Z is closed as required. ♠

Theorem 10.9.2 (Path Lifting Property) *Let $p : E \to B$ be a covering projection. Then given a path $\omega : I \to E$ and a point $e \in E$ such that, $p(e) = \omega(0)$, there exists a path $\overline{\omega} : I \to E$ such that, $p \circ \overline{\omega} = \omega$ and $\overline{\omega}(0) = e$.*

Proof: Let $Z = \{t \in I : \overline{\omega} \text{ is defined in } [0, t]\}$. Observe that by the very definition, Z is a sub-interval of I and contains 0. Let t_0 be the least upper bound of Z. It is enough to show that $t_0 \in Z$ and $t_0 = 1$.

Let V be an evenly covered open neighborhood of $\omega(t_0)$. For $0 < \epsilon < 1$ put

$$I_\epsilon = [t_0 - \epsilon, t_0 + \epsilon] \cap I.$$

Choose ϵ so that $\omega(I_\epsilon) \subset V$. Let U_i be the open neighborhood of $\overline{\omega}(t_0 - \epsilon/2)$, that is mapped homeomorphically onto V by p. Then $\lambda = p^{-1} \circ \omega$ is a lift of ω on I_ϵ. Observe that $\lambda(t_0 - \epsilon/2) = \overline{\omega}(t_0 - \epsilon/2)$. Therefore, by the uniqueness theorem, $\lambda(t) = \overline{\omega}(t), \forall\ t \in [t_0 - \epsilon/2, t_0 + \epsilon/2]$. Therefore the two lifts can be patched up. That is, $\overline{\omega}$ can be extended to a lift of ω on the interval $[0, t_0 + \epsilon] \cap I$. By the definition of t_0, we must then have $[0, t_0 + \epsilon] \cap I = I$ which means that $t_0 = 1$ and $t_0 \in Z$. ♠

Theorem 10.9.3 *Let $p : E \to B$ be a covering projection, $H : I \times I \to B$ be a homotopy of paths, $g : I \to E$ be a path such that, $p \circ g(t) = H(t, 0)$, $\forall\ t \in I$. Then there exists a continuous map $G : I \times I \to E$ such that $G(t, 0) = g(t), t \in I$ and $p \circ G = H$.*

Proof: By the Path Lifting Property of p as in the above theorem, it follows that there is a unique function $G : I \times I \to E$, such that, $p \circ G = H$, $G|I \times 0 = g$ and $G|t \times I$ is continuous for all $t \in I$. It remains to prove that, G is continuous as a function on $I \times I$.

As in the proof of proposition 7.3.1 in section 7.3, using Lebesgue covering lemma, we can subdivide $I \times I$ into finitely many squares $I_{k,l}$ such that $H(I_{k,l})$ is contained in an evenly covered open subset of B. We shall prove that $G|I_{k,l}$ is continuous for every k, l by induction on k.

For $k = 1$ and for each $I_{1,l}$, G is continuous on the bottom side L_l. Suppose $H(I_{1,l}) \subset V$, where V is evenly covered. By unique path lifting it follows that $G|L_1 = p_\alpha^{-1} \circ (H|L_l)$ where p_α^{-1} is the inverse of $p : V_\alpha \to V$ for some α. Therefore, $G(L_1) \subset V_\alpha$. Once again by uniqueness of the path lifting, it follows that $G|I_{1,l} = p_\alpha^{-1} \circ (H|I_{1,l})$. In particular, $G|I_{1,l}$ is continuous. Inductively, if we have proved the continuity of $G|I_{k-1,l}$ for all l then this will give us continuity of G on the bottom sides of $I_{k,l}$ for all l. Then repeating the above argument, it follows that $G|I_{k,l}$ are continuous for all l. This completes the proof of continuity of G. ♠.

Corollary 10.9.1 Let $\gamma_j, j = 0, 1$, be any two path homotopic paths with the same end points. Suppose $\tilde{\gamma}_j, j = 0, 1$, are their lifts with the same starting point. Then the two lifts have the same end points.

Proof: Let H be a path homotopy between γ_0 and γ_1 and let G be a lift of H with $G(t, 0) = \tilde{\gamma}_0$. Since $H(0, s) = \gamma_0(0) =: w_0$ for all s, and $H(1, s) = \gamma_0(1) =: w_1$, it follows that $G(0 \times I) \subset p^{-1} \subset p^{-1}(w_0)$ and $G(1 \times I) \subset p^{-1}(w_1$. Since $p^{-1}(w_j)$ are discrete subsets and $G(j \times I)$ are connected sets $j = 0, 1$, it follows that $G(0, 0) = G(0, s) = G(0, 1)$ and $G(1, 0) = G(1, s) = G(1, 1)$. Now $t \mapsto G(t, 1)$ is a lift of γ_1 at the point $G(0, 1) = \tilde{\gamma}_0(0) = \tilde{\gamma}_1(0)$. Therefore, by the uniqueness of the lift, $G(t, 1) = \tilde{\gamma}_1(t)$, for all $t \in I$. In particular, $\tilde{\gamma}_1(1) = G(1, 1) = G(1, 0) = \tilde{\gamma}_0(1)$. ♠

Theorem 10.9.4 *Let Y be a locally path connected and simply connected space and $f : Y \to B$ be any continuous map. Given $y_0 \in Y, w_0 \in \overline{X}$ such that $p(w_0) = f(y_0)$ there is a unique map $\tilde{f} : Y \to E$ such that $\tilde{f}(y_0) = w_0$ and $p \circ \tilde{f} = f$.*

Proof: For each point $y \in Y$, choose a path γ_y in Y joining y_0 to y and let $\tilde{\gamma}_y$ be the unique lift of $f \circ \gamma_y$ starting at w_0. Define $\tilde{f}(y) = \tilde{\gamma}_y(1)$, the end point. Clearly, $p \circ \tilde{f}(y) = f \circ \gamma_y(1) = f(y)$.

It remains to prove the continuity of $\tilde{f}$. Let $y \in Y$ be any point. Choose a path connected neighbourhood U of y in Y such that $f(U) \subset V$ where V is evenly covered. Suppose $\tilde{f}(y) \in V_\alpha$. Then on U we can consider the lift $p_\alpha^{-1} \circ f$ of f such that y is mapped to $\tilde{f}(y)$. Now for each $u \in U$, choose a path ω_u from y to u completely contained in U. Look at the path $\gamma_y * \omega_u$. The path $f(\gamma_y * \omega_u)$ gets lifted to $\tilde{\gamma}_y * \tilde{\omega}_u$ where $\tilde{\omega}_u = p_\alpha^{-1} \circ f \circ \omega_u$. Since Y is simply connected, the two paths γ_u and $\gamma_y * \omega_u$ are path-homotopic in Y. By the above theorem, it follows that the end points of their lifts are the same. Therefore $\tilde{f}(u) = p_\alpha^{-1} \circ f(u)$ on U. This proves the continuity of G. ♠

Corollary 10.9.2 Let Ω be a simply connected domain and $f : \Omega \to \mathbb{C} \setminus \{0, 1\}$ be any holomorphic mapping. Then there exists a holomorphic mapping $\tilde{f} : \Omega \to \boldsymbol{H}$ such that $\lambda \circ \tilde{f} = f$ where λ is the modular function.

Proof: Recall that λ is a covering projection. By the previous theorem, we have a continuous $\tilde{f}$ as required. Since λ is a local biholomorphic mapping, holomorphicity of $\tilde{f}$ follows. Indeed, in the last part of the proof of the above theorem, we obtain $\tilde{f}(u) = p_\alpha^{-1} \circ f(u)$ on U where p is now λ and f is holomorphic. ♠

We are now ready for Picard's theorem.

Proposition 10.9.1 Let $f : \mathbb{D}^* \to \mathbb{C} \setminus \{0, 1\}$ be a holomorphic mapping. Then 0 is a removable singularity or a pole of f.

Theorem 10.9.5 'Little' Picard Theorem *Let $f : \mathbb{C} \to \mathbb{C}$ be a non constant holomorphic function. Then f assumes all finite values except perhaps one.*

Proof: If f has a pole at infinity then we know that f is a (non constant) polynomial function and hence by Fundamental Theorem of Algebra, it assumes all finite values. On the other hand, if ∞ is an essential singularity, and if w_1 and w_2 are not in the image of f, then we can apply the above proposition to the function $\mu \circ f(1/z)$, where $\mu(z) = \frac{z - w_1}{w_2 - w_1}$. ♠.

Remark 10.9.3 Of course, Little Picard's theorem can also be arrived at by directly applying Liouville's theorem to a map $\tilde{f} : \mathbb{C} \to \boldsymbol{H}$ such that $\lambda \circ \tilde{f} = f$.

Theorem 10.9.6 'Big' Picard Theorem *Let $g : \mathbb{D}^* \to \mathbb{C}$ be a holomorphic mapping with 0 as an essential singularity. Then there exists at most one point $w \in \mathbb{C}$ such that in every punctured neighborhood of 0, f assumes all finite values except perhaps w.*

Proof: If possible, let w_1, w_2 be two distinct points in $\mathbb{C}$ and $0 < r < 1$ such that $f(\mathbb{D}_r^*) \subset \mathbb{C} \setminus \{w_1, w_2\}$. Consider the Möbius transformation $\mu(z) = \frac{z-w_1}{w_2-w_1}$ and put $f = \mu \circ g(rz)$ and apply the above proposition to arrive at a contradiction. ♠

Proof of the Proposition: We shall assume that 0 is an essential singularity and arrive at a contradiction. Consider the inverse of the Cayley map $\chi : \mathbb{D} \to \boldsymbol{H}$,

$$\chi^{-1}(w) = \frac{1+w}{1-w}$$

which is a biholomorphic mapping. Let $\lambda : \boldsymbol{H} \to \mathbb{C} \setminus \{0, 1\}$ be the modular function we have studied in the previous section. Let $\tilde{\lambda} = \lambda \circ \chi^{-1}$. Then $\tilde{\lambda} : \mathbb{D} \to \mathbb{C} \setminus \{0, 1\}$ is a covering projection. Let $U = \tilde{\lambda}(\mathbb{D}_{1/2})$. Apply Casorati-Weierstrass theorem to the function

$$z \mapsto f(e^{-2\pi} z)$$

to obtain a sequence z_n such that $|z_{n+1}| < |z_n| < 1,\ z_n \to 0$ and with the property $f(e^{-2\pi} z_n) \in U$. Define $g_n : \mathbb{D} \to \mathbb{C} \setminus \{0, 1\}$ by

$$g_n(z) = f(z_n e^{2\pi(\imath z - 1)}).$$

Choose $w_n \in \mathbb{D}_{1/2}$ such that $\tilde{\lambda}(w_n) = g_n(0)$. Let $\tilde{g}_n : \mathbb{D} \to \mathbb{D}$ be such that $\tilde{\lambda} \circ \tilde{g}_n = g_n$ and $\tilde{g}_n(0) = w_n$. By Montel's theorem 8.10.2 this family is normal. Therefore, by passing to a subsequence we may assume that $\tilde{g_n}$ converges uniformly on compact sets to a holomorphic function $\tilde{g} : \mathbb{D} \to \mathbb{C}$. Clearly $\tilde{g}(\mathbb{D}) \subset \overline{\mathbb{D}}$. If $\tilde{g}$ is non constant, then $\tilde{g}(\mathbb{D})$ will be an open subset of $\mathbb{C}$ and hence is actually contained in $\mathbb{D}$. If $\tilde{g}$ is a constant, then this constant will be equal to $\tilde{g}(0) = \lim_n \tilde{g_n}(0) = \lim_n w_n \in \overline{\mathbb{D}}_{1/2}$. Therefore, in either case, $\tilde{g}(\mathbb{D}) \subset \mathbb{D}$. This implies that g_n converges uniformly on compact sets to a function $g = \tilde{\lambda} \circ \tilde{g} : \mathbb{D} \to \mathbb{C} \setminus \{0, 1\}$. Choosing the compact set to be the closed interval $-1/2 \le \theta \le 1/2$, it follows that there exists $M > 0$ such that $|g_n(\theta)| < M$ for all n and $-1/2 \le \theta \le 1/2$. That is $|f(z_n e^{2\pi\imath\theta})| \le M$ for all n and $-1/2 \le \theta \le 1/2$. By maximum principle applied to f on the closed annulus $|z_{n+1}| \le |z| \le |z_n|$, it follows that $|f(z)| \le M$ in this annulus. Since $|z_n| \to 0$ this means f is bounded in $\mathbb{D}_r^*$ for $r = |z_1|$. Hence 0 is a removable singularity. ♠

Remark 10.9.4 It is possible to give a completely self-contained elementary proof of Montel's theorem based on Cantor's diagonal process without bringing in generalities we have discussed in chapter 8. However, contents of Chapter 8 have its own importance too.) Thus, the only two non trivial results used here are

(i) The modular function λ and

(ii) and the homotopy lifting property of covering projection,

neither of which may be covered in a first course in Complex Analysis.